

amplified by PCR, and subsequently used in a new DNA-templated synthesis round, which directly amplifies the recovered compounds.

SUMMARY OF THE INVENTION

5 The present invention combines the non-templated technique of Lerner with the templated technique of Walder and thereby provides an improved method for the generation of oligonucleotide-encoded libraries.

10 In a primary aspect the invention pertains to a method for synthesizing an encoded molecule comprising the steps of:

- a) Adding a linker molecule L to one or more reaction wells;
- b) Adding a molecule fragment to each of said reaction wells;
- c) Adding an oligonucleotide identifier to each of said reaction wells;
- 15 d) Subjecting said wells to conditions sufficient to allow said molecule fragments and said oligonucleotide identifiers to become attached to said linker molecule, or conditions sufficient for said molecule fragments to bind to other molecule fragments and sufficient for said oligonucleotide identifiers to bind to other oligonucleotide identifiers;

20 e) Combining the contents of said one or more reaction wells;

f) Optionally, distributing the combined product to one or more new reaction wells;

g) Optionally, repeating steps b) to e) one or more times;

Contacting the resulting bifunctional molecule(s) of step e) or g) with one or more (oligonucleotide) templates each capable of hybridizing to at least one of the

25 (oligonucleotide) templates added in step c); wherein

the linker molecule L contains at least one reactive group capable of reacting with a reactive group in the molecule fragments and at least one reactive group capable of reacting with a reactive group in the oligonucleotide;

30 the molecule fragments each contain at least one reactive group capable of reacting with a reactive group in the linker molecule L or a reactive group in another molecule fragment, and the reactive groups of each molecule fragment may be the same or different;

the oligonucleotide identifiers each contain at least one reactive group capable of reacting with a reactive group in the linker L or a reactive group in another oligonucleotide identifier, and the reactive groups of each oligonucleotide identifier may be the same or different;

35 the oligonucleotide identifier added to each well in step c) identifies the molecule fragment added to the same well in step b); the steps a) to d) may be performed in any order;

the steps b) to d) in step f) may also be performed in any order; the number of wells in steps a) and f) may be the same or different; the oligonucleotide template optionally is associated with a reactive group;

5 In another aspect, the invention relates to a method for identifying a molecule with desired characteristics, said method comprising synthesizing a library of encoded molecules by a method according to the invention.

10 In a further aspect, the invention pertains to a library of encoded molecules obtained or obtainable by a method according to the invention.

DEFINITIONS

As used herein, the term Bi-functional molecule means a bi-functional molecule consisting of an encoded molecule (e.g. a low molecular weight organic molecule) and an oligonucleotide (e.g. a single- or double-stranded DNA molecule), where the oligonucleotide sequence uniquely identifies the identity (structure) of the encoded molecule. The encoded molecule and the identifier are physically connected through a linker moiety. In certain embodiments, several oligonucleotides encode the same encoded molecule, or several encoded molecules are encoded by one oligonucleotide (see below under "Library of bi-functional molecules").

The bi-functional molecule can have one or more molecule fragments encoded by one or more oligonucleotide identifiers depending on the number of rounds of stage 1 split and mix synthesis used to generate the molecule.

25

Carrier molecule: Used interchangeably with carrier and bi-functional carrier molecule. A carrier molecule is a bi-functional molecule that is employed in a Stage 2 templated synthesis, and may be generated by e.g. stage 1 synthesis. It thus consists of an encoded molecule (made up of one or more molecule fragments) and an oligonucleotide identifier molecule (made up of one or more oligonucleotide identifiers) that uniquely identifies (encodes) the molecule fragment to which it is attached. The bi-functional carrier molecule can have one or more molecule fragments encoded by one or more oligonucleotide identifiers (depending on the number of rounds of stage 1 split and mix synthesis used to generate the carrier molecule).

35

Encoded molecule: The portion of the bi-functional molecule that is encoded by the oligonucleotide identifier of the bi-functional molecule. The encoded molecule is typically an organic molecule, typically of relatively low molecular weight compared to the oligonucleotide identifier to which it is attached. The encoded molecule may be released

from the identifier after its synthesis, to obtain the "free encoded molecule". The encoded molecule is typically attached to the identifier through a flexible linker.

Identifier: An oligonucleotide that encodes (specifies) the identity of the molecule fragment or encoded molecule to which it is attached. For the purpose of this invention, three kinds of identifiers are described:

Unit identifier, which is the oligonucleotide used in stage 1 synthesis to describe the identity of the molecule fragment that it becomes attached to, through the nascent bi-functional molecule, during a synthesis round in stage 1 synthesis.

10 Carrier identifier, which is the oligonucleotide component of a carrier molecule, i.e., the carrier identifier encodes the molecule fragment to which it is attached.

Template identifier, also termed identifier template, encodes the encoded molecule attached to it after templated synthesis, or in cases where no templated synthesis is performed prior to the screening of the bi-functional molecules, encodes and identifies the 15 encoded molecule attached to it.

Library of bi-functional molecules: A library of bi-functional molecules consists of a number of bi-functional molecules, each of which consists of an encoded molecule (e.g. low molecular weight organic molecules), attached to an identifier oligonucleotide (e.g. a 20 single- or double-stranded DNA molecule), where the oligonucleotide sequence uniquely identifies the identity (structure) of the encoded molecule to which it is attached. In certain embodiments, several oligonucleotides encode the same encoded molecule (i.e. several bi-functional molecules in the library carry the same encoded molecule but different oligonucleotide identifiers. In other embodiments, several different encoded molecules are 25 attached to the same oligonucleotide identifier.

Ligase enzyme: An enzyme that ligates together oligonucleotides. Ligase enzymes may also be non-protein-based catalysts that mediate the ligation of oligonucleotides, on single- or double-stranded form.

30 Ligation: A ligation reaction covalently links together molecules. It is primarily used here to describe the ligation of two oligonucleotides to produce one molecule, consisting of the two oligonucleotide sequences.

35 Linker L: The linker L is a molecule comprising a reactive group X, which is adapted for reaction with a molecule fragment, and a reactive group Z, which is adapted for ligation to an oligonucleotide fragment, and a linker Y, which connects X and Z.

Molecule fragment: A molecule fragment contains one or more reactive groups that may react with reactive groups of other molecule fragments.

Molecular Entity: Used interchangeably with encoded molecule.

5

Nucleic acid: Used interchangeably with oligonucleotide.

Nucleic acid analog : Used interchangeably with oligonucleotide analog and unnatural oligonucleotide.

10

Nucleotide: Nucleotides as used herein refers to both natural and unnatural nucleotides. Oligonucleotides made up of nucleotides are thus capable of sequence-specific hybridisation to natural oligonucleotides such as DNA and RNA. Nucleotides may differ from natural nucleotides by having a different phosphate moiety, sugar moiety, and/or 15 base moiety.

Oligonucleotide: Oligonucleotides comprise a number of nucleotides as defined above, i.e., oligonucleotides may comprise natural as well as unnatural nucleotides. Example oligonucleotides are thus DNA, RNA, PNA, morpholinos, and LNA, and may involve 20 unnatural bases as well.

Reactive group: Reactive groups are capable of reacting with other reactive groups to form a chemical bond. Reactive groups include $-NH_2$, $-COOH$, $-CHO$, $-OH$, $-NHR$, $-CSO_2OH$, phenylchloride, $-SH$, $-SS$, and many others. Example pairs of reactive groups, and the 25 resulting bonds formed, are shown in figure 6 and 7. The reaction between two reactive groups may be spontaneous under the conditions used, or can be catalyzed by enzymes, ribozymes or other organic or inorganic catalysts such as metals. Furthermore, additional reagents may be added that reacts with the reactive groups, in order to covalently link molecule fragments. The linkage between molecule fragments are thus typically of 30 covalent character. However, it may also be of non-covalent character. An example of such non-covalent bond between molecule fragments of an encoded molecule is the bond formed when adding a molecule fragment comprising a metal-chelate complex (e.g., $NTA-Zn^{++}$) to a nascent bi-functional molecule comprising an imidazole functionality.

35 Reactive units: Used herein interchangeably with reactive groups.

As used herein the term "well" defines a physical containment of reagents, molecule fragments etc. in a localized space. A "well" thus include the well of a microtiter plate, any container, a spot of a solution on a glass plate, or other type of solid support (microarray),

a reagent tube, a bead to which the reagents and molecules to be kept separated are attached, and any other type of well that separates different compositions of reagents, molecule fragments etc. as desired. The separation does not have to be absolute, but should preferably ensure that the major components of a given well are the desired 5 components. A nanocompartment where the molecule fragment to be attached to the nascent bi-functional molecule is held in the vicinity of the reactive group of the bi-functional molecule, by hybridisation of oligonucleotide strands, also is considered a "well", since the hybridisation of the oligonucleotides keeps one reactive group (e.g., of the incoming molecule fragment) in localised space as seen from the other reactive group 10 (e.g., of the bi-functional molecule). The complex of the bi-functional molecule and the incoming oligonucleotide-molecule fragment is therefore considered a nanocompartment and hence, under this invention, a "well".

The terms 'nucleic acid', 'nucleic acid molecule' and 'nucleic acid sequence' as used herein 15 refer to an oligomer or polymer of ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) or mimics/mimetics thereof. This term includes molecules composed of naturally-occurring nucleobases, sugars and covalent internucleoside (backbone) phosphodiester bond linkages as well as molecules having non-naturally occurring nucleobases, sugars and covalent internucleoside (backbone) linkages which function similarly or combinations 20 thereof. Such modified or substituted nucleic acids are often preferred over native forms because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for nucleic acid target and increased stability in the presence of nucleases and other enzymes, and are in the present context described by the terms "nucleic acid analogues" or "nucleic acid mimics". Preferred examples of nucleic acid mimics/mimetics 25 are peptide nucleic acid (PNA-), Locked Nucleic Acid (LNA-), xylo-LNA-, phosphorothioate-, 2'-methoxy-, 2'-methoxyethoxy-, morpholino- and phosphoramidate- containing molecules or the like.

The nucleic acid, nucleic acid molecule or nucleic acid sequence may, for instance, be 30 composed entirely of deoxyribonucleotides, entirely of ribonucleotides, entirely of nucleic acid mimics or analogues or chimeric mixtures thereof. The monomers are typically linked by internucleotide phosphodiester bond linkages. Nucleic acids typically range in size from a few monomeric units, e.g., 5-40, when they are commonly referred to as oligonucleotides, to several thousands of monomeric units. Whenever a nucleic acid or a 35 nucleic acid sequence is represented, it will be understood that the nucleotides are in 5' to 3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, and "T" denotes thymidine, unless otherwise noted.

In the present context 'complementary sequence' or 'complement' refers to nucleotide sequences which will anneal to a nucleic acid molecule of the invention under stringent conditions. The term "stringent conditions" refers to general conditions of high, weak or low stringency. The term "stringency" is well known in the art and is used in reference to 5 the conditions (temperature, ionic strength and the presence of other compounds such as organic solvents) under which nucleic acid hybridisations are conducted. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences, as compared to conditions of "weak" or "low" stringency.

As an example, high stringency hybridisation conditions comprise (1) low ionic strength and high temperature for washing, such as 0.015 M NaCl/0.0015 M sodium citrate, pH 7.0 (0.1xSSC) with 0.1% sodium dodecyl sulfate (SDS) at 50°C; (2) hybridisation in 50% (vol/vol) formamide with 5 x Denhardt's solution (0.1% (wt/vol) polyvinylpyrrolidone), 50 mM sodium 15 serum albumin/0.1% (wt/vol) Ficoll/0.1% (wt/vol) polyvinylpyrrolidone), 50 mM sodium phosphate buffer at pH 6.5 and 5 x SSC at 42°C; or (3) hybridisation in 50% formamide, 5 x SSC, 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5 x Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42°C with washes at 42°C in 0.2 x SSC and 0.1% SDS.

When referring to 'identical sequences' herein, again it is meant sequences having a certain degree of sequence identity. The sequences may thus be from 1-100%, such as at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, 25 at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 99% or 100% identical.

The term 'sequence identity' indicates a quantitative measure of the degree of homology between two nucleic acid sequences of equal length. If the two sequences to be compared 30 are not of equal length they must be aligned to give the best possible fit, allowing the insertion of gaps or, alternatively, truncation at the ends of the polypeptide sequences or nucleotide sequences. The sequence identity can be calculated as $\frac{(N_{id}-N_{del})/N_{id}}{N_{id}}$, wherein N_{id} is the total number of non-identical residues in the two sequences when aligned and wherein N_{del} is the number of residues in one of the sequences. Hence, the DNA sequence 35 AGTCAGTC will have a sequence identity of 75% with the sequence AATCAATC ($N_{id}=2$ and $N_{del}=8$). A gap is counted as non-identity of the specific residue(s), i.e. the DNA sequence AGTCGTC will have a sequence identity of 75% with the DNA sequence AGTCAGTC ($N_{id}=2$ and $N_{del}=8$).

In all polypeptide or amino acid based embodiments of the invention the percentage of sequence identity between one or more sequences is based on alignment of the respective sequences as performed by clustalW software (<http://www.ebi.ac.uk/clustalW/index.html>) 5 using the default settings of the program. With respect to the nucleotide-based embodiments of the invention, the percentage of sequence identity between one or more sequences is also based on alignments using the clustalW software with default settings. For nucleotide sequence alignments these settings are: Alignment=3Dfull, Gap Open 10.00, Gap Ext. 0.20, Gap separation Dist. 4, DNA weight matrix: identity (IUB).

10

In the present context, by 'amplification reaction' is meant a reaction that produces one or more copies of a sequence of nucleic acids by repeated extension of a probe or primer.

'Extension' may occur by virtue of polymerisation of individual nucleotide monomers, as in 15 PCR, or it may occur by the addition of prefabricated oligonucleotide segments, as in LCR, or by a combination of these as in gap LCR or Repair Chain Reaction (RCR). Though not essential to the invention, ideally the extension reactions are performed repeatedly and the extension products themselves may serve as templates to produce an exponential generation of products.

20

' T_m ' can be defined as the temperature at which 50% of a nucleic acid and its perfect complement are in duplex. The denaturation of double stranded nucleic acids causes a shift in the absorbance of UV light at 260 nm wavelength, an effect which can be assayed by determining the optical density at 260 nm (OD_{260}). T_m is defined as the 25 temperature corresponding to 50% denaturation, i.e. where the (OD_{260}) is midway between the value expected for double stranded nucleic acids and the value expected for single stranded nucleic acids. The T_m of perfectly complementary duplexes can be calculated as follows:

$$\begin{aligned} \text{DNA: } T_m &= 81.5 + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\% \text{GC}) - 500/\text{length} \\ 30 \text{ RNA: RNA-DNA: } T_m &= 79.8 + 18.5(\log_{10}[\text{Na}^+]) + 0.58(\% \text{GC}) + 11.8(\% \text{GC})^2 - 820/\text{length} \end{aligned}$$

$$\text{Oligonucleotides: } T_m = 2(\text{no. of AT pairs}) + 4(\text{no. of GC pairs})$$

As used herein, 'nucleic acid analogue' is understood to mean a structural analogue of DNA 35 or RNA, designed to hybridise to complementary nucleic acid sequences (1). Through modification of the internucleotide linkage(s), the sugar, and/or the nucleobase, nucleic acid analogues may attain any or all of the following desired properties: 1) optimised hybridisation specificity or affinity, 2) nuclease resistance, 3) chemical stability, 4)

solubility, 5) membrane-permeability, and 6) ease or low costs of synthesis and purification. Examples of nucleic acid analogues include, but are not limited to, peptide nucleic acids (PNA), locked nucleic acids "LNA", 2'-O-methyl nucleic acids, 2'-fluoro nucleic acids, phosphorothioates, and metal phosphonates. Nucleic acid analogues are described in 5 (2) and (3).

ABBREVIATIONS

L = Linker molecule.

10

$R_{1(1-m)}$ = Molecule fragments from repertoire 1. The repertoire used in the corresponding round thus comprises m different molecule fragments.

$R_{2(1-m)}$ = Molecule fragments from repertoire 2. The repertoire used in the corresponding 15 round thus comprises m different molecule fragments.

$O_{1(1-m)}$ = Oligonucleotides coding for molecule fragments of repertoire 1. m different oligonucleotide sequences are thus used to encode the m different molecule fragments.

$O_{2(1-m)}$ = Oligonucleotides coding for molecule fragments of repertoire 2. m different oligonucleotide sequences are thus used to encode the m different molecule fragments.

$R_{1,1}$ thus is molecule fragment no. 1 of repertoire 1; $R_{1,2}$ is molecule fragment no. 2 of repertoire 2; $R_{2,17}$ thus is molecule fragment no. 17 of repertoire 2; etc.

25

Oligo $O_{1,1}$ codes for molecule fragment $R_{1,1}$; oligo $O_{1,2}$ codes for molecule fragment $R_{1,2}$; oligo $O_{2,m}$ codes for molecule fragment $R_{2,m}$ etc.

30 DETAILED DESCRIPTION OF THE INVENTION

The invention provides a method for combining the advantages of encoded molecule fragments made by split and mix synthesis with the advantages of template directed synthesis of molecules. The method provided in the invention further has the advantage 35 that molecules in the library after selection in a screening assay can be easily identified and amplified for use in subsequent screening procedures.

In outline, the invention combines methods for producing encoded molecule fragments with methods for template directed synthesis of molecules. In the first step unique

10

Identifiers are attached to different molecule fragments, so that each different molecule fragment is "encoded" by a unique identifier. In the next step, a second molecule fragment and thereafter a second identifier are attached. Like in the first step the identifier encodes the newly attached molecule fragment. This process can be continued until the desired number and diversity of encoded molecules have been made. The resulting molecules are termed carrier molecules, where a given carrier molecule contains organic molecule fragments that are linked together, and where the linked molecule fragments are also linked to the identifiers that have also been linked together.

10 These carrier molecules are hereafter brought together two-and-two by complementary binding of the identifier molecules to a template that directs which two carrier molecules are brought together. Thus, each template binds two carriers. In certain embodiments, the number of carriers that bind to the same template may be 3, 4, 5, 6, 7 or more. The proximity of the molecule fragments, once they are juxtaposed on the template, allows for high fidelity chemical transfer reactions to occur in which one molecule fragment is transferred to the other and linked. In this way, the encoded molecule fragments produced by the first methods can be linked in all possible combinations to create very large libraries of molecules. Since each molecule is defined by a template the identity of the molecule, after it for example has been selected in a screening assay, can easily be determined.

20 Furthermore, the templates of the selected molecules can easily be amplified and used for amplification of the selected molecule fragments, which can then be subjected to further rounds of screening and selection. Each round of screening will thus enrich the pool of selected molecules with molecules having the highest affinity or best function in the screening assay.

25

In a first aspect the invention pertains to a method for synthesizing an encoded molecule comprising the steps of:

- a) Adding a linker molecule L to one or more reaction wells;
- b) Adding a molecule fragment to each of said reaction wells;
- 30 c) Adding an oligonucleotide identifier to each of said reaction wells;
- d) Subjecting said wells to conditions sufficient to allow said molecule fragments and said oligonucleotide identifiers to become attached to said linker molecule, or conditions sufficient for said molecule fragments to bind to other molecule fragments and sufficient for said oligonucleotide identifiers to bind to other oligonucleotide identifiers;
- 35 e) Combining the contents of said one or more reaction wells;
- f) Optionally, distributing the combined product to one or more new reaction wells;
- g) Optionally, repeating steps b) to e) one or more times;

11

h) Contacting the resulting bifunctional molecule(s) of step e) or g) with one or more (oligonucleotide) templates each capable of hybridizing to at least one of the oligonucleotide identifiers added in step c); wherein the linker molecule L contains at least one reactive group capable of reacting with a reactive group in the molecule fragment and at least one reactive group capable of reacting with a reactive group in the oligonucleotide; the molecule fragments each contain at least one reactive group capable of reacting with a reactive group in the linker molecule L or a reactive group in another molecule fragment, and the reactive groups of each molecule fragment may be the same or different;

10 the oligonucleotide identifiers each contain at least one reactive group capable of reacting with a reactive group in the linker L or a reactive group in another oligonucleotide identifier, and the reactive groups of each oligonucleotide identifier may be the same or different;

the oligonucleotide identifier added to each well in step c) identifies the molecule fragment added to the same well in step b);

15 the steps a) to d) may be performed in any order; the steps b) to d) in step f) may also be performed in any order; the number of wells in steps a) and f) may be the same or different; the oligonucleotide template optionally is associated with a reactive group.

20

The process outlined above may be seen as a combination of a step 1 synthesis and a step 2 synthesis, wherein the step 1 synthesis comprises the steps a) to g) and the step 2 synthesis is carried out in step h.

25 As mentioned, steps b) to e) may be repeated one or more times. It is to be understood, however, that when step d) is performed for the first time the conditions must allow for the molecule fragments and said oligonucleotide identifiers to become attached to said linker molecule. Accordingly, in the first round of step 1 synthesis, the conditions must be so that the molecule fragments do not only bind to other molecule fragments.

30

The present invention thus provides a method for the fast generation of very large libraries of DNA-encoded molecules in an amplifiable system. In the invention the synthesis of a library is carried out in two stages by combination of two different methods.

In stage 1 a repertoire of different bi-functional molecules are generated. These bi-functional molecules are called "bi-functional carrier molecules", "carrier molecules" or simply "carriers". Each bi-functional molecule consists of a molecular entity and a unique identifier that codes for the molecular entity. In one embodiment of the invention the molecular entity is a polypeptide and the identifier encoding the polypeptide is a DNA oligonucleotide.

In stage 2, the carrier molecules produced in stage 1 are used in a templated synthesis reaction, in order to create very large libraries of molecule fragments each attached to a unique identifier and each with the ability of being amplified. In one embodiment the templates are DNA oligonucleotides that are complementary to the oligonucleotide

5 identifiers of the bi-functional carrier molecules produced in stage 1.

Stage 1:

In principle any method could be used for producing the bi-functional carrier molecules. In one embodiment this is done by split and mix synthesis of the molecular entities and

10 Identifier molecules. In a further embodiment this is done using the method in the example described below and illustrated in figure 1.

Example of split and mix synthesis of carriers:

The following example of split and mix synthesis of carriers is a modification of the method described by Lerner et al (Lerner, R. et al. (1993), *European Patent Specification, "EP 0643778B1"*).

The example is outlined in figure 1. In round 1, a linker molecule is first added to wells in a microtitre plate. Repertoires of different amino acids (R, 1 through m) are hereafter added

20 to the wells, one type of amino acid per well (i.e., a specific amino acid to each well), and operatively linked to the linker molecule. Other types of molecule fragments can be used as well, but in the present example amino acids are used. Furthermore, a unique identifier (O, 1 through m), here a DNA oligonucleotide, is added to each well and operatively linked to the linker molecule. Each well now contains a bi-functional molecule that consists of a

25 linker molecule linked to an amino acid and an identifier oligonucleotide. Each well has a different amino acid and each amino acid is linked, via the linker, to a unique DNA oligo. The sequence of the oligo encodes the type of amino acid added to that well. The oligonucleotide can be of any length depending on the number of different amino acids that must be encoded. In the present example a DNA oligo of 12 nucleotides is used.

30 The content of all the wells are hereafter pooled, and then split into wells on a new microtitre plate. Each well on the new plate will now contain all the different species of bi-functional molecules generated by the above reactions. A new round of synthesis (round 2) that is similar to round 1 can hereafter be applied: first an amino acid is added to each

35 well, one species of amino acid for each well, and linked to the amino group of the amino linker of the bi-functional molecules in the well (instead of linking to the amino group of the linker as in round 1). Likewise, different DNA oligos are added to each well and linked to the DNA-portion of the bi-functional molecules in the wells. Like in the first round, the oligos in each well have different sequences so that each newly added amino acid is

uniquely encoded by the sequence of the newly added oligo. Each well now contains a bi-functional molecule consisting of a di-peptide (two amino acids) linked to an oligonucleotide through a linker molecule. The oligonucleotide, which in the present example is now 24 nucleotides long, uniquely identifies the di-peptide to which it is linked.

5 The sequence of the first 12 nucleotides encodes the species of the first amino acid in the di-peptide and the sequence of the next 12 nucleotides encodes the identity of the second amino acid.

The amino acids used in the second round can be from the same repertoire as used in round 1, or it can be a different set of amino acids or other completely different types of

10 molecule fragments. In the present example they are from a different repertoire of amino acids in order to increase the diversity of the molecules.

The desired coupling reactions of the reactive groups are ensured by protecting and de-protecting the relevant reactive groups of the molecules during the process.

15 The content of each well is hereafter pooled as after round 1. A new round of synthesis can hereafter be applied by splitting the pooled bi-functional molecules into wells on a new plate and repeating round 2, which would create bi-functional molecules consisting of tri-peptides (or other kinds of compounds made up of three molecule fragments) attached to 36-mer identifier oligos through a linker. Principally the molecules could be increased in

20 size and diversity by applying new rounds of synthesis until the desired compound is obtained. In the present example the synthesis is terminated after round 2. The pool of bi-functional molecules that is obtained will have a diversity that depends on the number of amino acids (or other molecule fragments) used for the synthesis in round 1 and 2. If 1000 different molecule fragments are used in both rounds the number of different bi-functional

25 molecules in the pool will be one million (1000 times 1000). The length of the attached oligos (which is in this example 24 nucleotides) is more than enough to uniquely encode the one million different compounds formed.

Thus, using this split and mix approach a repertoire of bi-functional carrier molecules have

30 been generated. This repertoire of bi-functional carrier molecules are then used in Stage 2 to generate, through a DNA-templating approach, an even bigger repertoire of bi-functional molecules.

Stage 2:

35 The next stage of the library synthesis uses the bi-functional carrier molecules generated in stage 1 in DNA templated synthesis reactions, which essentially links together the bi-functional carrier molecules provided by stage 1 in different combinations. The molecules generated this way are all uniquely identifiable by attached templates that encode the bi-functional carrier molecules that were combined by the reaction. One advantage of the

templated carrier synthesis reaction is that the template, in addition to being a unique identifier that encodes the carrier molecules that are linked together by the reaction, also brings the molecules into close proximity and thereby essentially eliminates the likelihood of reactions to occur with other template-carrier complexes. Another advantage, important for the present invention, is that the DNA template can easily be amplified (e.g. by PCR using appropriate primer binding sites included in the template) and serve for the amplification of molecules that are isolated in a screening assay, even if only a small number of each molecule is present in the isolated samples. The amplification allows for the execution of further rounds of screening and selection, until molecules with desired characteristics (e.g., affinity, specificity, or catalytic activity) are identified.

Several different methods of templated synthesis have been proposed as described in the background of the invention. In principle any of these methods could be used to carry out the reaction set in stage 2. In one embodiment of the invention the templated synthesis is done using the method described below and illustrated in figure 2.

Example of template directed synthesis using bi-functional carrier molecules generated in Stage 1 above:

The following example is a modification of the method described by Walder et al (Walder, J.A. et al. (1979) *Proc. Natl. Acad. Sci.*, **76**; 51-55). Figure 2 shows an example of templated synthesis using templates that are complementary to pairs of bi-functional carrier molecules generated in step 1. Thus, two bi-functional carrier molecules, each comprising an encoded molecule generated in Stage 1 above (thus, in this example, a di-peptide) that is linked to a unique DNA oligonucleotide identifier (in this example, a 24-nucleotide DNA oligo), are brought into close proximity through hybridisation of the carriers' DNA portion to a complementary DNA template. The sequence of the DNA template thus determines which carrier molecules bind to the template. A reaction between the reactive groups of the two carriers hybridised to the same template may now be induced. In the present example, the design of the molecule fragments in Stage 1 led to carrier molecules that can react through an acylation reaction in Stage 2. Thus, an acyl transfer reaction (see e.g., figure 3 and 6A) leads to the transfer of the encoded molecule (here, a di-peptide) of one carrier onto the encoded molecule (here, a di-peptide) of the other carrier, resulting in the generation of a bi-functional molecule where the encoded molecule (here, a tetra-peptide) is linked, covalently or non-covalently (in this example non-covalently) to a template that encodes the combination of the di-peptides and thus, ultimately encodes the tetrapeptide.

The acyl transfer leaves the donor carrier as an "empty" oligonucleotide without encoded molecule. Both the empty donor carrier and the acceptor carrier molecule (now carrying the full encoded molecule, here the tetrapeptide) may be attached to the DNA template

throughout the acyl transfer reaction. In order to keep the encoded molecule (tetrapeptide) physically associated with a DNA that encodes it, the two carriers must be linked together, or alternatively the carrier that carries the encoded molecule (the tetrapeptide) must be linked covalently or non-covalently to the DNA template. Several ways to ensure the association of the encoded molecule with a DNA molecule that encodes it can be envisioned, some of which are shown in the section "alternative methods for templated synthesis using carriers".

In a simple embodiment, the link to the DNA template is kept by conducting the subsequent screening of the library under conditions that does not disrupt the hybridisation of the oligonucleotide identifiers of the carriers to the template.

In the present invention, as shown in the example in figure 2, the carrier molecules used in the templated synthesis in stage 2 originate from the repertoire of different bi-functional di-peptide carrier molecules generated by the reactions in stage 1. These are di-peptides attached through a linker to a 24-mer oligonucleotide, where the first 12 nucleotides encode the first amino acid of the di-peptide and the next 12 nucleotides encode the second amino acid. A library of DNA templates is therefore synthesised where each template consists of a coding sequence of 48 nucleotides; the first 24 nucleotides are complementary to the 24-mer identifier oligonucleotide of one di-peptide carrier molecule from stage 1 and the last 24 nucleotides are complementary to another 24-mer oligo from the repertoire of carrier molecules. Thus, the sequence of the template encodes which di-peptide carriers can bind and thereby it encodes and uniquely identifies the four amino acid polypeptides that result from the transfer of one di-peptide onto the other in the acyl transfer reaction. The two carrier molecules from stage 1 that are encoded by the template can be from the same repertoire or from two different sets of bi-functional carriers. In the present example they originate from different repertoires.

If one million different bi-functional carrier molecules were generated in stage 1 in the example and combined in stage 2 with another set of one million different bi-functional carriers, the diversity of the resulting library of molecules generated in stage 2 would be 10^{12} (1 million times 1 million) different four amino acid polypeptides. If different repertoires of amino acids and carriers are used in stage 1 and 2, respectively, a total of up to 4000 different amino acids would be required to carry out the synthesis as in the present example. If the same repertoire of amino acids was used only 1000 different amino acids would be required, but the resulting library would be less complex although it would still contain 10^{12} different molecules.

During Stage 1, the oligonucleotide identifiers may be linked to the linker or to the oligonucleotide portion of the nascent bi-functional molecule by enzymatic means, e.g. by

ligases (e.g. T4 DNA ligase, E. coli DNA ligase, or T7 DNA ligase for double stranded DNA fragments, or T4 RNA ligase for single-stranded DNA fragments), or by chemical ligation. Several methods for chemical ligation are known in the art, such as the 5'-phosphorimidazolide method (Visscher, J., Shwartz, A.W. *Journal of Molecular Evolution* 5 (1988), 28, 3-6; Zhao, Y., Thorson, J.S. *Journal of Organic Chemistry* (1998), 63, 7568-7572), or the 3'-phosphothioate method (Alvarez et al., *Journal of Organic Chemistry* (1999), 64, 6319-28; Pirrung et al., *Journal of Organic Chemistry* (1998) 63, 241-46).

Other means of ligating together two oligonucleotides include the use of CNBr as a condensing agent for chemical ligation (Sokolova, N.I., et al., *FEBS Letters* (1988), 232, 153-155; Dolinnaya, N.G. et al., *Nucl. Acids. Res.* (1993), 21, 5403-5407); reductive amination between juxtaposed amine and aldehyde groups (Goodwin, J.T., and Lynn, D.G., *J. Am. Chem. Soc.* (1992), 114, 9197-9198); disulfide bond formation (e.g. reaction of a thiol and an activated disulfide such as pyridyl disulfide); reaction between pyrophosphate-activated 5'-phosphate and 3'-hydroxyl, to form a phosphodiester bond (Rohatgi, R., et al., 15 *J. Am. Chem. Soc.* (1996), 118, 3332-3339); a tosyl displacement reaction (Herriehin, M.K., et al, *J. Am. Chem. Soc.* (1995), 117, 10151-10152); and reaction of 5'-iodonucleoside and 3-phosphothioate (Yanzheng, X., and Kool, E.T., *Tetrahedron Letters* (1997), 38, 5595-5598).

Some of the ligation reactions generate natural phosphodiester bonds, while other ligation reactions generate non-natural bonds between the ligated oligonucleotides. Some of the more effective methods for forming natural phosphodiester bonds utilise activation of a phosphoryl group with CNBr, cyanimidazole or water soluble carbodilimide, described in the following papers: Wang, E., Yanagawa, H. *Biochemistry* (1986), 25, 7423-7430; Shabarova, Z.A., *Biochemie* (1988), 70, 1323-1334; Dolinnaya, N.G., Merenkova, I.N., 25 Shabarova, Z.A. *Nucleosides Nucleotides* (1994), 13, 2169-2183; Kool, E.T. *J. Am. Chem. Soc.* (1991), 113, 6265-6266; Ashley, G.W., Kushlan, D.M. *Biochemistry* (1991), 30, 2927-2933; Luebke, K.J., Dervan, P.B., *J. Am. Chem. Soc.* (1991), 113, 7447-7448; Luebke, K.J., Dervan, P.B. *Nucl. Acids Res.* (1992), 20, 3005-3009; Gao, H., et al., *Bioconjugate Chem.* (1994), 5, 445-453. In particular, 3'-OH and 5'-phosphate groups, or 30 functional derivatives thereof, have been used to ligate together oligonucleotides, to form a natural phosphodiester bond. Other examples of reactions that may be used to link together identifiers, and the bonds resulting from those reactions, are shown in figures 6 and 7.

The standard phosphoramidite method for oligonucleotide synthesis may also be applied for chemical ligation. For both enzymatic and chemical ligation it is preferable that the DNA fragments to be ligated are on double-stranded form and with overhangs. In stage 1 synthesis, the identifier may be attached as a double stranded oligonucleotide or a single stranded oligonucleotide, and likewise, the identifier of the nascent bi-functional molecule may be on double- or single-stranded form. The incoming oligonucleotide may be attached

by ligation (chemical or enzymatic) of one or two strands, or may be attached by non-covalent interaction of the incoming identifier and the oligonucleotide of the bi-functional molecule, e.g. by annealing of complementary oligonucleotide regions of the incoming oligonucleotide and the bi-functional molecule, or by a third oligonucleotide and the incoming identifier/bi-functional molecule.

In a preferred embodiment the oligonucleotide identifiers are not de-protected (after the phosphoramidite synthesis of individual identifiers), prior to their linking in the Stage 1 synthesis. This will allow conditions to be applied in the linking of molecule fragments and linking of oligonucleotide identifiers during the stage 1 process that could not otherwise easily have been applied, such as the use of certain organic solvents. As an example, if the reaction that is used for linking two unit identifiers together during a stage 1 synthesis require relatively strong acidic conditions, and the use of an organic solvent such as DMF, it may be desirable to use oligonucleotides, for which the nucleic acid bases have not been deprotected, as this will make them less vulnerable to acidic conditions, and also, will 15 make them more soluble in organic solvents such as DMF. Then, after the stage 1 synthesis of the carrier molecule, the oligonucleotide component may be deprotected, to allow efficient sequence specific interaction between the carrier and the template in a stage 2 synthesis.

Finally, the ligation of unit identifiers may lead to linear as well as branched products.

An alternative method for performing stage 1 synthesis involves the hybridisation between the bi-functional molecule and the incoming oligonucleotide, where then an extension reaction is performed using e.g. a polymerase to extend from the 3'-end of the bi-functional molecule, to add an identifier sequence (Freskgård et al., WO 2004/039825 A2, 25 Enzymatic Encoding). Polymerases that may be used in this approach include Reverse transcriptase, DNA polymerase, RNA polymerase, Taq polymerase, Pfu polymerase, Vent polymerase, Klenow fragment, and many others.

During templated synthesis in stage 2 the oligonucleotide portions of the carrier molecules may be ligated together by enzymatic or chemical ligation.

During both stage 1 and stage 2, the bi-functional molecules may be immobilised to solid support, allowing simple and efficient removal of reagents and by-products, as well as allowing the use of organic solvents that might otherwise precipitate molecule fragments, oligonucleotides or bi-functional molecules.

A wide range of conditions may be applied to mediate the reaction between the nascent bi-functional molecule and the molecule fragment. These include the addition of catalysts, base, acid, or reagents that take part in the reaction. The latter may react with both the nascent bi-functional molecule and the molecule fragment, and thus may end up as a

bridge or linker that links the nascent bi-functional molecule and the molecule fragment. Such linking moieties include di-carboxylic acids (where for example the reactive groups to become linked are two amines), or di-amines (where two carboxylic acids become linked through the linking moiety).

5 Reactive groups of the present invention include aldehydes, hydroxyls, isocyanate, thiols, amines, esters, thioesters, carboxylic acids, triple bonds, double bonds, ethers, acid chlorides, phosphates, imidazoles, halogenated aromatic rings, any precursors thereof, or any protected reactive groups. Examples of reactive groups that can be employed during stage 1, and the bonds resulting from these reactions, are shown in figures 6 and 7.

10 Reactions that may be employed during stage 1 synthesis include acylation (formation of amide, pyrazolone, isoxazolone, pyrimidine, comarine, quinolinon, phthalhydrazide, diketopiperazine, benzodiazepinone, and hydantoin), alkylation, vinylation, disulfide formation, Wittig reaction, Horner-Wittig-Emmans reaction, arylation (formation of biaryl or vinylarene), condensation reactions, cycloadditions ([2+4], [3+2]), addition to carbon-carbon multiple bonds, cycloaddition to multiple bonds, addition to carbon-hetero multiple bonds, nucleophilic aromatic substitution, transition metal catalyzed reactions, as well as the reaction types listed below, and may involve formation of ethers, thioethers, secondary amines, tertiary amines, beta-hydroxy ethers, beta-hydroxy thioethers, beta-hydroxy amines, beta-amino ethers, amides, thioamides, oximes, sulfonamides, di- and tri- compounds, alkyn substituted aromatic compounds, vinyl substituted aromatic compounds, alkyn substituted aromatic compounds, biaryl compounds, hydrazines, hydroxylamine ethers, substituted cycloalkenes, substituted cycloalkenes, substituted 1, 2, 3 triazoles, substituted cycloalkenes, beta-hydroxy ketones, beta-hydroxy aldehydes, vinyl ketones, vinyl aldehydes, substituted alkenes, substituted alkenes, substituted amines, and many others.

Molecule fragments, identifiers and reagents.

In stage 1 synthesis, any number of molecule fragments may be linked to the nascent bi-functional molecule during a synthesis round. Thus, 0, 1, 2, 3 or more molecule fragments may be linked to the nascent bi-functional molecule in a given synthesis round. In the same synthesis round an identifier encoding the molecule fragment(s) that were linked to the nascent bi-functional molecule is linked to the nascent oligonucleotide of the bi-functional molecule. Thus, one identifier may encode a combination of several molecule fragments, and also, may encode molecule fragment(s), reagent(s), catalyst(s) and conditions employed. In such cases, two different identifiers may encode the same molecule fragment but different reaction conditions. If desired, more than one identifier may be linked to the nascent oligonucleotide. For example, if two molecule fragments are reacted with the nascent bi-functional molecule, one may choose to also link two identifiers

to the nascent oligonucleotide. However, in most cases one identifier is used to encode all the molecule fragments added in a given synthesis round.

In a given synthesis round, the attachment of the molecule fragment(s) to the nascent bi-functional molecule can be performed before, simultaneously with, or after the attachment of the identifier(s) encoding said molecule fragments.

In stage 2 synthesis, any number of carriers may be added simultaneously, in order to transfer 0, 1, 2, 3, 4 or more molecule fragments to the nascent bi-functional molecule. Alternatively, carriers are added sequentially or partly sequentially (e.g. two carriers at a time), and therefore, the molecule fragments of the final bi-functional molecule are made up of molecule fragments that have been transferred simultaneously, partly sequentially or fully sequentially (i.e., one at a time).

During the templated synthesis of stage 2, the molecule fragments are attached to the oligonucleotide identifiers prior to reaction between carriers. During the templated synthesis, the molecule fragment of one carrier is transferred to another carrier. This may involve a direct transfer in which the reaction between two reactive groups directly leads to a transfer of one molecule fragment from one carrier to another (figure 3A). Reactions that mediate the direct transfer of one molecule fragment from one carrier to another include the reactions listed in figure 6. Alternatively, a transfer may involve first a reaction between the reactive groups, followed by cleavage of the bond linking one molecule fragment to the carrier, which will result in the transfer of that molecule fragment onto the other carrier (figure 3B). Types of direct and indirect transfer reactions, as well as the bonds resulting from the reaction of reactive groups, are shown in figures 6 and 7. A number of reactions for stage 1 and stage 2 synthesis are listed below.

In certain embodiments the molecule fragments remain associated with both carriers after the reaction step of Stage 2, i.e., it is not a requirement in the present invention that the molecule fragments are transferred from one carrier to another.

Reaction conditions compatible with stage 1 and stage 2 synthesis of bifunctional molecules.

Stage 1 and stage 2 reactions can occur in aqueous or non-aqueous (i.e., organic) solutions, or a mixture of one or more aqueous and non-aqueous solutions. In aqueous solutions, reactions can be performed at pH ranges from about 2 to about 12, or preferably from about 2 to about 10, or more preferably from about 4 to about 10. The reactions used in DNA-templated chemistry preferably should not require very basic conditions (e.g., pH > 10), or very acidic conditions (e.g., pH < 1, pH < 2, pH < 4), because extreme conditions may lead to degradation or modification of the nucleic acid template and/or encoded molecule being synthesized. The aqueous solution can

contain one or more inorganic salts, including, but not limited to, NaCl, Na₂SO₄, KCl, Mg²⁺, Mn²⁺, etc., at various concentrations.

Organic solvents suitable for stage 1 and stage 2 reactions include, but are not limited to, methylene chloride, chloroform, dimethylformamide, and organic alcohols, including methanol and ethanol. To permit quantitative dissolution of reaction components in organic solvents, quaternized ammonium salts, such as, for example, long chain tetraalkylammonium salts, can be added (Jost *et al.* (1989) NUCLEIC ACIDS RES. 17:2143; Melnikov *et al.* (1999) LANGMUIR 15: 1923-1928).

Stage 1 or stage 2 reactions may require a catalyst, such as, for example, homogeneous, heterogeneous, phase transfer, and asymmetric catalysis. In other embodiments, a catalyst is not required. The presence of additional, accessory reagents not linked to a nucleic acid are preferred in some embodiments. Useful accessory reagents can include, for example, oxidizing agents (e.g., NaIO₄; reducing agents (e.g., NaCNBH₃); activating reagents (e.g., EDC, NHS, and sulfo-NHS); transition metals such as nickel (e.g., Ni(NO₃)₂), rhodium (e.g., RhCl₃), ruthenium (e.g., RuCl₃), copper (e.g., Cu(NO₃)₂), cobalt (e.g., CoCl₂), iron (e.g., Fe(NO₃)₃), osmium (e.g., OsO₄), titanium (e.g., TiCl₄ or titanium tetraisopropoxide), palladium (e.g., NaPdCl₄), or Ln; transition metal ligands (e.g., phosphines, amines, and halides); Lewis acids; and Lewis bases.

Reaction conditions preferably are optimized to suit the nature of the reactive units and oligonucleotides used.

20

Reaction types compatible with stage 1 and stage 2 synthesis.

Known chemical reactions for synthesizing polymers, small molecules, or other chemical compounds can be used in stage 1 and stage 2 synthesis reactions. Thus, reactions such as those listed in *March's Advanced Organic Chemistry, Organic Reactions, Organic*

Syntheses, organic text books, Journals such as *Journal of the American Chemical Society, Journal of Organic Chemistry, Tetrahedron*, etc., and Carruther's *Some Modern Methods of Organic Chemistry* can be used. The chosen reactions preferably are compatible with nucleic acids such as DNA or RNA or are compatible with the modified nucleic acids used as the template.

Reactions useful in stage 1 and stage 2 synthesis include, for example, substitution reactions, carbon-carbon bond forming reactions, elimination reactions, acylation reactions, and addition reactions. An illustrative but not exhaustive list of aliphatic nucleophilic substitution reactions useful in the present invention includes, for example, S_N2 reactions, S_N1 reactions, S_Ni reactions, allylic rearrangements, nucleophilic substitution at an aliphatic trigonal carbon, and nucleophilic substitution at a vinylic carbon.

35

Specific aliphatic nucleophilic substitution reactions with oxygen nucleophiles include, for example, hydrolysis of alkyl halides, hydrolysis of gen-dihalides, hydrolysis of 1,1,1-trihalides, hydrolysis of alkyl esters or inorganic acids, hydrolysis of diazo ketones, hydrolysis of acetal and enol ethers, hydrolysis of epoxides, hydrolysis of acyl halides, hydrolysis of anhydrides, hydrolysis of carboxylic esters, hydrolysis of amides, alkylation with alkyl halides (Williamson Reaction), epoxide formation, alkylation with inorganic esters, alkylation with diazo compounds, dehydration of alcohols, transesterification, alcoholysis of epoxides, alkylation with onium salts, hydroxylation of silanes, alcoholysis of acyl halides, alcoholysis of anhydrides, esterification of carboxylic acids, alcoholysis of carboxylic esters (transesterification), alcoholysis of amides, alkylation of carboxylic acid salts, cleavage of ether with acetic anhydride, alkylation of carboxylic acids with diazo compounds, acylation of carboxylic acids with acyl halides; acylation of carboxylic acids with carboxylic acids, formation of oxonium salts, preparation of peroxides and hydroperoxides, preparation of inorganic esters (e.g., nitrites, nitrates, sulfonates), preparation of alcohols from amines, and preparation of mixed organic-inorganic anhydrides.

Specific aliphatic nucleophilic substitution reactions with sulfur nucleophiles, which tend to be better nucleophiles than their oxygen analogs, include, for example, attack by SH at an alkyl carbon to form thiols, attack by S at an alkyl carbon to form thioethers, attack by SH or SR at an acyl carbon, formation of disulfides, formation of Bunte salts, alkylation of sulfuric acid salts, and formation of alkyl thiocyanates.

20

Aliphatic nucleophilic substitution reactions with nitrogen nucleophiles include, for example, alkylation of amines, N-arylation of amines, replacement of a hydroxy by an amino group, transamination, transamidation, alkylation of amines with diazo compounds, amination of epoxides, amination of oxetanes, amination of aziridines, amination of alkanes, formation of isocyanides, acylation of amines by acyl halides, acylation of amines by anhydrides, acylation of amines by carboxylic acids, acylation of amines by carboxylic esters, acylation of amines by amides, acylation of amines by other acid derivatives, N-alkylation or N-arylation of amides and imides, N-acylation of amides and imides, formation of aziridines from epoxides, formation of nitro compounds, formation of azides, formation of isocyanates and isothiocyanates, and formation of azoxy compounds.

Aliphatic nucleophilic substitution reactions with halogen nucleophiles include, for example, attack at an alkyl carbon, halide exchange, formation of alkyl halides from esters of sulfuric and sulfonic acids, formation of alkyl halides from alcohols, formation of alkyl halides from ethers, formation of haloethers from epoxides, cleavage of carboxylic esters with lithium iodide, conversion of diazo ketones to alpha-halo ketones, conversion of amines to halides, conversion of tertiary amines to cyanamides (the von Braun reaction), formation of acyl halides from carboxylic acids, and formation of acyl halides from acid derivatives.

35

Aliphatic nucleophilic substitution reactions using hydrogen as a nucleophile include, for example, reduction of alkyl halides, reduction of tosylates, other sulfonates, and similar compounds,

hydrogenolysis of alcohols, hydrogenolysis of esters (Barton-McCombie reaction), hydrogenolysis of nitriles, replacement of alkoxyl by hydrogen, reduction of epoxides, reductive cleavage of carboxylic esters, reduction of a C-N bond, desulfurization, reduction of acyl halides, reduction of carboxylic acids, esters, and anhydrides to aldehydes, and reduction of amides to aldehydes.

5

Although certain carbon nucleophiles may be too nucleophilic and/or basic to be used in certain embodiments of the invention, aliphatic nucleophilic substitution reactions using carbon nucleophiles include, for example, coupling with silanes, coupling of alkyl halides (the Wurtz reaction), the reaction of alkyl halides and sulfonate esters with Group I (I A), 10 and II (II A) organometallic reagents, reaction of alkyl halides and sulfonate esters with organocuprates, reaction of alkyl halides and sulfonate esters with other organometallic reagents; allylic and propargylic coupling with a halide substrate, coupling of organometallic reagents with esters of sulfuric and sulfonic acids, sulfoxides, and sulfones, coupling involving alcohols, coupling of organometallic reagents with carboxylic esters,

15 coupling of organometallic reagents with compounds containing an ester linkage, reaction of organometallic reagents with epoxides, reaction of organometallics with aziridine, alkylation at a carbon bearing an active hydrogen, alkylation of ketones, nitriles, and carboxylic esters, alkylation of carboxylic acid salts, alkylation at a position alpha to a heteroatom (alkylation of 1,3-dithianes), alkylation of dihydro-1,3-oxazine (the Meyers

20 synthesis of aldehydes, ketones, and carboxylic acids), alkylation with trialkylboranes, alkylation at an alkynyl carbon, preparation of nitriles, direct conversion of alkyl halides to aldehydes and ketones, conversion of alkyl halides, alcohols, or alkanes to carboxylic acids and their derivatives, the conversion of acyl halides to ketones with organometallic compounds, the conversion of anhydrides, carboxylic esters, or amides to ketones with

25 organometallic compounds, the coupling of acyl halides, acylation at a carbon bearing an active hydrogen, acylation of carboxylic esters by carboxylic esters (the Claisen and Dieckmann condensation), acylation of ketones and nitriles with carboxylic esters, acylation of carboxylic acid salts, preparation of acyl cyanides, and preparation of diazo ketones, ketonic decarboxylation.

30 Reactions which involve nucleophilic attack at a sulfonyl sulfur atom may also be used in the present invention and include, for example, hydrolysis of sulfonic acid derivatives (attack by OH), formation of sulfonic esters (attack by OR), formation of sulfonamides (attack by nitrogen), formation of sulfonyl halides (attack by halides), reduction of sulfonyl chlorides (attack by hydrogen), and preparation of sulfones (attack by carbon).

35 Aromatic electrophilic substitution reactions may also be used in stage 1 and stage 2 synthesis schemes. Hydrogen exchange reactions are examples of aromatic electrophilic substitution

reactions that use hydrogen as the electrophile. Aromatic electrophilic substitution, reactions which use nitrogen electrophiles include, for example, nitration and nitro-dehydrogenation, nitrosation of nitroso-de-hydrogenation, diazonium coupling, direct introduction of the diazonium group, and amination or amino-dehydrogenation. Reactions of this type with

5 sulfur electrophiles include, for example, sulfonation, sulfo-dehydrogenation, halosulfonation, halosulfo-dehydrogenation, sulfurtization, and sulfonylation. Reactions using halogen electrophiles include, for example, halogenation, and halo-dehydrogenation. Aromatic electrophilic substitution reactions with carbon electrophiles include, for example, Friedel-Crafts alkylation, alkylation, alkylation, Friedel-Crafts arylation (the Scholl reaction), Friedel-Crafts acylation, formylation 10 with disubstituted formamides, formylation with zinc cyanide and HCl (the Gatterman reaction), formylation with chloroform (the Reimer-Tiemann reaction), other formylations, formyl-dehydrogenation, carboxylation with carbonyl halides, carboxylation with carbon dioxide (the Kolbe-Schmitt reaction), amidation with isocyanates, *N*-alkylcarbamoyl-dehydrogenation,

15 haloalkylation, hydroxyalkyl-dehydrogenation, cyclodehydration of aldehydes and ketones, dialkylamino-dehydrogenation, thioalkylation, acylation with nitriles (the Hoesch reaction), cyanation, and cyano-de hydrogenation. Reactions using oxygen electrophiles include, for example, hydroxylation and hydroxy-dehydrogenation.

Rearrangement reactions include, for example, the Fries rearrangement, migration of a nitro group, 20 migration of a nitroso group (the Fischer-Hepp Rearrangement), migration of an arylazo group, migration of a halogen (the Orton rearrangement), migration of an alkyl group, etc. Other reaction on an aromatic ring include the reversal of a Friedel-Crafts alkylation, decarboxylation of aromatic aldehydes, decarboxylation of aromatic acids, the Jacobsen reaction, deoxygenation, desulfonation, hydro-desulfonation, dehalogenation, hydro-dehalogenation, and hydrolysis of organometallic 25 compounds.

Aliphatic electrophilic substitution reactions are also useful. Reactions using the S_N1 , S_N2 (front), S_N2 (back), S_E , addition-elimination, and cyclic mechanisms can be used in the present invention.

Reactions of this type with hydrogen as the leaving group include, for example, hydrogen exchange (deuterio-de-hydrogenation, deuteration), migration of a double bond, and keto-enol 30 tautomerization. Reactions with halogen electrophiles include, for example, halogenation of aldehydes and ketones, halogenation of carboxylic acids and acyl halides, and halogenation of sulfoxides and sulfones. Reactions with nitrogen electrophiles include, for example, aliphatic diazonium coupling, nitrosation at a carbon bearing an active hydrogen, direct formation of diazo compounds, conversion of amides to alpha-azido amides, direct amination at an activated position, 35 and insertion by nitriles. Reactions with sulfur or selenium electrophiles include, for example, sulfonylation, sulfonation, and selenylation of ketones and carboxylic esters. Reactions with carbon electrophiles include, for example, acylation at an aliphatic carbon, conversion of aldehydes to beta-keto esters or ketones, cyanation, cyano-de-hydrogenation, alkylation of alkanes, the Stork

enamine reaction, and insertion by carbenes. Reactions with metal electrophiles include, for example, metalation with organometallic compounds, metalation with metals and strong bases, and conversion of enolates to silyl enol ethers. Aliphatic electrophilic substitution reactions with metals as leaving groups include, for example, replacement of metals by hydrogen, reactions between organometallic reagents and oxygen, reactions between organometallic reagents and peroxides, oxidation of trialkylboranes to borates, conversion of Grignard reagents to sulfur compounds, halo-demetalation, the conversion of organometallic compounds to amines, the conversion of organometallic compounds to ketones, aldehydes, carboxylic esters and amides, cyano-de-metalation, transmetalation with a metal, transmetalation with a metal halide, transmetalation with an organometallic compound, reduction of alkyl halides, metallo-de-halogenation, replacement of a halogen by a metal from an organometallic compound, decarboxylation of aliphatic acids, cleavage of alkoxides, replacement of a carboxyl group by an acyl group, basic cleavage of beta-keto esters and beta-diketones, haloform reaction, cleavage of non-enolizable ketones, the Haller-Bauer reaction,

15 cleavage of alkanes, decyanation, and hydro-de-cyanation. Electrophilic substitution reactions at nitrogen include, for example, diazotization, conversion of hydrazines to azides, *N*-nitrosation, *N*-nitroso-de-hydrogenation, conversion of amines to azo compounds, *N*-halogenation, *N*-halo-de-hydrogenation, reactions of amines with carbon monoxide, and reactions of amines with carbon dioxide.

20 Aromatic nucleophilic substitution reactions may also be used in the present invention. Reactions proceeding via the S_NAr mechanism, the S_N1 mechanism, the benzyl mechanism, the S_{W1} mechanism, or other mechanism, for example, can be used. Aromatic nucleophilic substitution reactions with oxygen nucleophiles include, for example, hydroxy-de-halogenation, alkali fusion of sulfonate salts, and replacement of OR or OR. Reactions with sulfur nucleophiles include, for example, replacement by SH or SR. Reactions using nitrogen nucleophiles include, for example, replacement by NH_2 , NHR , or NR_2 , and replacement of a hydroxy group by an amino group: Reactions with halogen nucleophiles include, for example, the introduction of halogens. Aromatic nucleophilic substitution reactions with hydrogen as the nucleophile include, for example, reduction of phenols and phenolic esters and ethers, and reduction of halides and nitro compounds. Reactions with carbon nucleophiles include, for example, the Rosenmund-von Braun reaction, coupling of organometallic compounds with aryl halides, ethers, and carboxylic esters, arylation at a carbon containing an active hydrogen, conversions of aryl substrates to carboxylic acids, their derivatives, aldehydes, and ketones, and the Ullmann reaction. Reactions with hydrogen as the leaving group include, for example, alkylation, arylation, and amination of nitrogen heterocycles. Reactions with N_2^+ as the leaving group include, for example, hydroxy-de-diazoliation, replacement by sulfur-containing groups, iodo-de-diazoliation, and the Schlemann reaction. Rearrangement reactions include, for example, the von Richter rearrangement, the Sommelet-Hauser rearrangement, rearrangement of aryl hydroxylamines, and the Smiles rearrangement. Reactions involving free radicals can also be used, although the free radical

reactions used in nucleotide-templated chemistry should be carefully chosen to avoid modification or cleavage of the nucleotide template. With that limitation, free radical substitution reactions can be used in the present invention. Particular free radical substitution reactions include, for example, substitution by halogen, halogenation at an alkyl carbon, allylic halogenation, benzylic halogenation, 5 halogenation of aldehydes, hydroxylation at an aliphatic carbon, hydroxylation at an aromatic carbon, oxidation of aldehydes to carboxylic acids, formation of cyclic ethers, formation of hydroperoxides, formation of peroxides, acyloxylation, acyloxy-de-hydrogenation, chlorosulfonation, nitration of alkanes, direct conversion of aldehydes to amides, amidation and amination at an alkyl carbon, simple coupling at a susceptible position, coupling of alkynes, 10 arylation of aromatic compounds by diazonium salts, arylation of activated alkenes by diazonium salts (the Meerwein arylation), arylation and alkylation of alkenes by vinyltin compounds (the Stille reaction), arylation and alkylation of aromatic compounds by peroxides, photochemical arylation of aromatic compounds, alkylation, acylation, and carbalkoxylation of nitrogen heterocycles. Particular reactions in which N_2^+ is the leaving group include, for example, replacement of the diazonium group by hydrogen, replacement of the diazonium group by chlorine or bromine, nitro-de-diazoliation, replacement of the diazonium group by sulfur-containing groups, aryl dimerization with diazonium salts, methylation of diazonium salts, vinylation of diazonium salts, arylation of diazonium salts, and conversion of diazonium salts to aldehydes, ketones, or carboxylic acids.

20 Free radical substitution reactions with metals as leaving groups include, for example, coupling of Grignard reagents, coupling of boranes, and coupling of other organometallic reagents. Reaction with halogen as the leaving group are included. Other free radical substitution reactions with various leaving groups include, for example, desulfurization with Raney Nickel, conversion of sulfides to organolithium compounds, decarboxylative dimerization (the Kolbe reaction), the Hunsdiecker reaction, decarboxylative alkylation, and decarbonylation of aldehydes and acyl halides. Reactions involving additions to carbon-carbon multiple bonds are also used in the stage 1 and stage 2 synthesis schemes. Any mechanism may be used in the addition reaction including, for example, electrophilic addition, nucleophilic addition, free radical addition, and cyclic mechanisms. 30 Reactions involving additions to conjugated systems can also be used. Addition to cyclopropane rings can also be utilized. Particular reactions include, for example, isomerization, addition of hydrogen halides, hydration of double bonds, hydration of triple bonds, addition of alcohols, addition of carboxylic acids, addition of H_2S and thiols, addition of ammonia and amines, addition of amides, addition of hydrazoic acid, hydrogenation of double and triple bonds, other reduction of double and triple bonds, reduction of the double and triple bonds of conjugated systems, hydrogenation of aromatic rings, reductive cleavage of cyclopropanes, hydroboration, other hydrometalations, addition of alkanes, addition of alkenes and/or alkynes to alkenes and/or alkynes (e.g., pi-cation cyclization reactions, hydro-alkenyl-addition), ene reactions, the Michael reaction, addition of organometallics to double and triple bonds not conjugated to carbonyls, the addition of

two alkyl groups to an alkyne, 1,4-addition of organometallic compounds to activated double bonds, addition of boranes to activated double bonds, addition of tin and mercury hydrides to activated double bonds, acylation of activated double bonds and of triple bonds, addition of alcohols, amines, carboxylic esters, aldehydes, etc., carbonylation of double and triple bonds,

- 5 hydrocarboxylation, hydroformylation, addition of aldehydes, addition of HCN, addition of silanes, radical addition, radical cyclization, halogenation of double and triple bonds (addition of halogen, halogen), halolactonization, halolactamization, addition of hypohalous acids and hypohalites (addition of halogen, oxygen), addition of sulfur compounds (addition of halogen, sulfur), addition of halogen and an amino group (addition of halogen, nitrogen), addition of NOX and NO₂X (addition of halogen, nitrogen), addition of XN₃ (addition of acyl halides (addition of halogen, carbon), hydroxylation (addition of oxygen, oxygen) (*e.g.*, asymmetric dihydroxylation reaction with OSO₄), dihydroxylation of aromatic rings, epoxidation (addition of oxygen, oxygen) (*e.g.*, Sharpless asymmetric epoxidation), 15 photooxidation of dienes (addition of oxygen, oxygen), hydroxysulfenylation (addition of oxygen, sulfur), oxyamination (addition of oxygen, nitrogen), diamination (addition of nitrogen, nitrogen), formation of aziridines (addition of nitrogen), aminosulfenylation (addition of nitrogen, sulfur), acylacyloxylation and acylamidation (addition of oxygen, carbon or nitrogen, carbon), 1,3-dipolar addition; (addition of oxygen, nitrogen, carbon), 20 Diels-Alder reaction, heteroatom Diels-Alder reaction, all carbon 3 + 2 cycloadditions, dimerization of alkenes, the addition of carbenes and carbenoids to double and triple bonds, trimerization and tetramerization of alkynes, and other cycloaddition reactions.

In addition to reactions involving additions to carbon-carbon multiple bonds, addition reactions to carbon-hetero multiple bonds can be used in nucleotide-templated chemistry.

- 25 Exemplary reactions include, for example, the addition of water to aldehydes and ketones (formation of hydrates), hydrolysis of carbon-nitrogen double bond, hydrolysis of aliphatic nitro compounds, hydrolysis of nitriles, addition of alcohols and thiols to aldehydes and ketones, reductive alkylation of alcohols, addition of alcohols to isocyanates, alcoholysis of nitriles, formation of xanthates, addition of H₂S and thiols to carbonyl compounds, 30 formation of bisulfite addition products, addition of amines to aldehydes and ketones, addition of amides to aldehydes, reductive alkylation of ammonia or amines, the Mannich reaction, the addition of amines to isocyanates, addition of ammonia or amines to nitriles, addition of amines to carbon disulfide and carbon dioxide, addition of hydrazine derivative to carbonyl compounds, formation of oximes, conversion of aldehydes to nitriles, formation of gem-dihalides from aldehydes and ketones, reduction of aldehydes and ketones to alcohols, reduction of the carbon-nitrogen double bond, reduction of nitriles to amines, 35 reduction of nitriles to aldehydes, addition of Grignard reagents and organolithium reagents to aldehydes and ketones, addition of other organometallics to aldehydes and ketones, addition of trialkylsilyl silanes to aldehydes and ketones, addition of conjugated alkenes to

aldehydes (the Baylis-Billmeyer reaction), the Reformatsky reaction, the conversion of carboxylic acid salts to ketones with organometallic compounds, the addition of Grignard reagents to acid derivatives, the addition of organometallic compounds to CO₂ and CS₂, addition of organometallic compounds to C=N compounds, addition of carbenes and

- 5 diazoalkanes to C=N compounds, addition of Grignard reagents to nitriles and isocyanates, the Aldol reaction, Mukaiyama Aldol and related reactions, Aldol-type reactions between carboxylic esters or amides and aldehydes or ketones, the Knoevenagel reaction (*e.g.*, the Nef reaction, the Favorskii reaction), the Peterson alkenylation reaction, the addition of active hydrogen compounds to CO₂ and CS₂, the Perkin reaction, Darzens glycidic ester condensation, the Tollens reaction, the Wittig reaction, the Tebbe alkenylation, the Petasis alkenylation, alternative alkenylations, the Thorpe reaction, the Thorpe-Ziegler reaction, addition of silanes, formation of cyanohydrins, addition of HCN to C=N and C-N bonds, the Prins reaction, the benzoin condensation, addition of radicals to C=O, C=S, C=N compounds, the Ritter reaction, acylation of aldehydes and ketones, addition of aldehydes to aldehydes, the addition of isocyanates to isocyanates (formation of carbodilimides), the conversion of carboxylic acid salts to nitriles, the formation of epoxides from aldehydes and ketones, the formation of episulfides and episulfones, the formation of beta-lactones and oxetanes (*e.g.*, the Paterno-Buchi reaction), the formation of beta-lactams, etc. Reactions involving addition to isocyanides include the addition of water to isocyanides, the Passerini reaction, the Ugi reaction, and the formation of metalated aldimines.
- 20 Elimination reactions, including alpha, beta, and gamma eliminations, as well as extrusion reactions, can be performed using nucleotide-templated chemistry, although the strength of the reagents and conditions employed should be considered. Preferred elimination reactions include reactions that go by E1, E2, E1cB, or E2C mechanisms. Exemplary reactions include, for example, reactions in which hydrogen is removed from one side (*e.g.*, dehydration of alcohols, cleavage of ethers to alkenes, the Chugaev reaction, ester decomposition, cleavage of quaternary ammonium hydroxides, cleavage of quaternary ammonium salts with strong bases, cleavage of amine oxides, pyrolysis of keto-ylids, decomposition of toluene-p-sulfonylhydrazones, cleavage of amine oxides, cleavage of selenoxides, cleavage of sulfones, dehydrohalogenation of alkyl halides, dehydrohalogenation of acyl halides, dehydrohalogenation of sulfonyl halides, elimination of boranes, conversion of alkenes to alkynes, decarbonylation of acyl halides, reactions in which neither leaving atom is hydrogen (*e.g.*, deoxygenation of vicinal diols, cleavage of cyclic thionocarbonates, conversion of epoxides to episulfides and alkenes, the Ramberg-Bäcklund reaction, conversion of aziridines to alkenes, dehalogenation of vicinal dihalides, dehalogenation of alpha-halo acyl halides, and elimination of a halogen and a hetero group), fragmentation reactions (i.e., reactions in which carbon is the positive leaving group or the electrofuge, such as, for example, fragmentation of gamma-amino and gamma-hydroxy halides, fragmentation of 1,3-diols, decarboxylation of beta-hydroxy carboxylic acids, decarboxylation of (3-lactones, fragmentation of alpha-beta-epoxy hydrazones, elimination of CO from bridged bicyclic compounds, and elimination of CO₂ from bridged bicyclic compounds), reactions in which C=N or C=C bonds are

formed (e.g., dehydration of aldoximes or similar compounds, conversion of ketoximes to nitriles, dehydration of unsubstituted amides, and conversion of N-alkylformamides to isocyanides), reactions in which C=O bonds are formed (e.g., pyrolysis of beta-hydroxy alkenes), and reactions in which N=N bonds are formed (e.g., eliminations to give diazoalkenes). Extrusion reactions

5 Include, for example, extrusion of N₂ from pyrazolines, extrusion of N₂ from pyrazoles, extrusion of N₂ from triazolines, extrusion of CO, extrusion of CO₂, extrusion of SO₂, the Storr synthesis, and alkene synthesis by twofold extrusion.

Rearrangements, including, for example, nucleophilic rearrangements; electrophilic

rearrangements, prototropic rearrangements, and free-radical rearrangements, can also be

10 performed using stage 1 and stage 2 synthesis schemes. Both 1,2 rearrangements and non-1,2

rearrangements can be performed. Exemplary reactions include, for example, carbon-to-carbon

migrations of R, H, and Ar (e.g., Wagner-Meerwein and related reactions, the Pinacol

rearrangement, ring expansion reactions, ring contraction reactions, acid-catalyzed

rearrangements of aldehydes and ketones, the dienone-phenol rearrangement, the Favorskii

15 rearrangement, the Arndt-Eistert synthesis, homologation of aldehydes, and homologation of

ketones), carbon-to-carbon migrations of other groups (e.g., migrations of halogen, hydroxyl,

amino, etc.; migration of boron; and the Neber rearrangement), carbon-to-nitrogen migrations of

R and Ar (e.g., the Hofmann rearrangement, the Curtius rearrangement, the Lossen

rearrangement, the Schmidt reaction, the Beckmann rearrangement, the Stieglitz rearrangement,

20 and related rearrangements), carbon-to-oxygen migrations of R and Ar (e.g., the Baeyer-Villiger

rearrangement and rearrangement of hydroperoxides), nitrogen-to-carbon, oxygen-to-carbon, and

sulfur-to-carbon migration (e.g., the Stevens rearrangement, and the Wittig rearrangement),

boron-to-carbon migrations (e.g., conversion of boranes to alcohols (primary or otherwise),

conversion of boranes to aldehydes, conversion of boranes to carboxylic acids, conversion of vinylic

25 boranes to alkenes, formation of alkynes from boranes and acetylides, formation of alkenes from

boranes and acetylides, and formation of ketones from boranes and acetylides), electrocyclic

rearrangements (e.g., of cyclobutenes and 1,3-cyclohexadienes, or conversion of stilbenes to

phenanthrenes), sigmatropic rearrangements (e.g., (1,1) sigmatropic migrations of hydrogen,

(1,3) sigmatropic migrations of carbon, conversion of vinylcyclopropanes to cyclopentenes, the

30 Cope rearrangement, the Claisen rearrangement, the Fischer indole synthesis, (2,3) sigmatropic

rearrangements, and the benzidine rearrangement), other cyclic rearrangements (e.g., metathesis

of alkenes, the di-n-methane and related rearrangements, and the Hofmann-Löffler and related

reactions), and non-cyclic rearrangements (e.g., hydride shifts, the Chapman rearrangement, the

Wallach rearrangement, and dyotropic rearrangements).

35 Oxidative and reductive reactions may also be performed using stage 1 and stage 2 synthesis schemes. Exemplary reactions may involve, for example, direct electron transfer, hydride transfer, hydrogen-atom transfer, formation of ester intermediates, displacement mechanisms, or addition-elimination mechanisms. Exemplary oxidations include, for example, eliminations of hydrogen

(e.g., aromatization of six-membered rings, dehydrogenations yielding carbon-carbon double bonds, oxidation or dehydrogenation of alcohols to aldehydes and ketones, oxidation of phenols and aromatic amines to quinones, oxidative cleavage of ketones, oxidative cleavage of aldehydes, oxidative cleavage of alcohols, ozonolysis, oxidative cleavage of double bonds and aromatic rings,

5 oxidation of aromatic side chains, oxidative decarboxylation, and bisdecarboxylation), reactions

Involving replacement of hydrogen by oxygen (e.g., oxidation of methylene to carbonyl, oxidation

of methylene to OH, CO₂R, or OR, oxidation of arylmethanes, oxidation of ethers to carboxylic esters

and related reactions, oxidation of aromatic hydrocarbons to quinones, oxidation of amines or nitro

compounds to aldehydes, ketones, or dihalides, oxidation of primary alcohols to carboxylic acids or

10 carboxylic esters, oxidation of alkenes to aldehydes or ketones, oxidation of amines to nitroso

compounds and hydroxylamines, oxidation of primary amines, oximes, azides, isocyanates, or

nitroso compounds, to nitro compounds, oxidation of thiols and other sulfur compounds to sulfonic

acids), reactions in which oxygen is added to the substrate (e.g., oxidation of alkynes to alpha-

diketones, oxidation of tertiary amines to amine oxides, oxidation of thioesters to sulfoxides and

15 sulfones, and oxidation of carboxylic acids to peroxy acids, and oxidative coupling reactions (e.g.,

coupling involving carbanions, dimerization of silyl enol ethers or of lithium enolates, and oxidation

of thiols to disulfides).

Exemplary reductive reactions include, for example, reactions involving replacement of oxygen by

hydrogen (e.g., reduction of carbonyl to methylene in aldehydes and ketones; reduction of

20 carboxylic acids to alcohols; reduction of amides to amines; reduction of carboxylic esters to

ethers; reduction of cyclic anhydrides to lactones and acid derivatives to alcohols; reduction of

carboxylic esters to alcohols; reduction of carboxylic acids and esters to alkanes, complete

reduction of epoxides; reduction of nitro compounds to amines; reduction of nitro compounds to

hydroxylamines; reduction of nitroso compounds and hydroxylamines to amines; reduction of

25 oximes to primary amines or aziridines; reduction of azides to primary amines; reduction of

nitrogen compounds, and reduction of sulfonyl halides and sulfonic acids to thiols), removal of

oxygen from the substrate (e.g., reduction of amine oxides and azoxy compounds; reduction of

sulfoxides and sulfones; reduction of hydroperoxides and peroxides; and reduction of aliphatic nitro

compounds to oximes or nitriles), reductions that include cleavage (e.g., de-alkylation of amines

30 and amides; reduction of azo, azoxy, and hydrazo compounds to amines; and reduction of

disulfides to thiols), reductive coupling reactions (e.g., bimolecular reduction of aldehydes and

ketones to 1,2-diols, bimolecular reduction of aldehydes or ketones to alkenes, acylin ester

condensation, reduction of nitro to azoxy compounds; and reduction of nitro to azo compounds),

and, reductions in which an organic substrate is both oxidized and reduced (e.g., the Cannizzaro

35 reaction, the Tishchenko reaction, the Pummerer rearrangement, and the Willgerodt reaction).

Examples of cleavable linkers/protecting groups that may be cleaved in order to transfer molecule fragments as described above, or may be used to protect and de-protect functional groups during the synthesis of the library of bi-functional molecules, and

conditions mediating the cleavage of these linkers/protecting groups, are shown in figure 8. Cleavable linkers can be cleaved in any number of ways, e.g., by photolysis or increased temperature, or by the addition of acid, base, enzymes, ribozymes, other catalysts, or any other agents.

5

The linkers of the present invention may for example be chosen from the following list: Carbohydrates and substituted carbohydrates, polyvinyl, acetylene or polyacetylene, aryl/hetaryl and substituted aryl/hetaryl, ethers and polyethers such as e.g. polyethylenglycol and substituted polyethers, amines, polyamines and substituted

10 polyamines, single- or double-stranded oligonucleotides, and polyamides and natural and unnatural polypeptides.

To maintain a physical link between the identifier and the encoded molecule (in the case of stage 2 synthesis, the template and the encoded molecule), at least one non-cleavable 15 linker is needed. The non-cleavable linker may of course be cleavable under certain conditions, but is non-cleavable under the conditions that lead to the bi-functional molecule employed in the screening. This non-cleavable linker is preferably flexible, enabling it to expose the encoded molecule in an optimal way. Preferably the length of the flexible linker is in the range of 1-50 Å, more preferably 5-30 Å, most preferably 10-25 Å. 20 Preferably the linker is both flexible and inert; polyethylene glycol (PEG) is an appropriate linker.

Under certain conditions it may be desirable to be able to cleave the linker after the screening of the library of bi-functional molecules has been done, for example in order to perform a mass spectrometric analysis of the encoded molecule without the identifier attached, or to 25 perform other types of assays on the free encoded molecule.

In an alternative embodiment, the linker contains an oligonucleotide moiety which may serve as an annealing site for an oligonucleotide that carries a reagent, catalyst or molecule fragment. The annealing of the reagent-, catalyst- or molecule fragment- 30 oligonucleotide will serve to provide the reagent, catalyst or molecule fragment in a high local concentration, thereby improving the efficiency of the desired reaction.

Alternative methods for templated synthesis using bi-functional carrier molecules:

Alternative methods (to the example shown in figure 2) can be envisioned for carrying out 35 the template directed reactions set in stage 2 of the invention. The alternative methods provide different ways of bringing together and reacting the bi-functional carrier molecules provided by stage 1. Some examples of such alternative methods are shown in figure 4. Any of these methods could be used to carry out stage 2 of the invention.

Example 1 (see example 1, figure 4):

This is identical to the example provided in figure 2, except that the carriers are ligated together before reaction. X and Y represent the two different repertoires of carrier molecules from stage 1 used in the example of figure 2. Once the oligonucleotide

5 Identifiers of the carriers have annealed to the template, they are ligated to each other and the acyl transfer reaction is hereafter carried out (the acyl transfer could also be carried out prior to the ligation step). The ligation step may be performed by a ligase, or may be performed in absence of an enzyme.

10 *Example 2 (see example 2, figure 4):*

In example 2, the molecule fragments are attached to the ends of the oligonucleotide identifiers that are distal to each other after annealing to the template. After ligation of the oligonucleotide identifiers of the carriers, the molecule fragments are brought in contact for reaction by denaturing the duplex DNA (ligated carriers annealed to template) and 15 removal of the template (in order to have increased flexibility of the now single stranded DNA sequence that carries the molecule fragments). The correct juxtaposition of the molecule fragments can be further ensured by including complementary sequences in the ends of the oligonucleotide identifiers that are proximal to the molecule fragments before transfer.

20 After reaction of the molecule fragments the single stranded DNA can be made double stranded by a standard extension reaction using a DNA polymerase and a primer annealing to the end of the DNA.

25 *Example 3 (see example 3, figure 4):*

Example 3 shows how the oligonucleotide identifier of the carrier that carries the combined molecule fragments can be ligated to the template. This is achieved by including a region of self-complementary sequence on the template that after annealing will juxtapose the ends of the carrier oligo and the template.

30 As in example 2 it may also in this example be desirable to perform an extension reaction, using a primer that anneals to one of the ends of the DNA strand that is attached to the encoded molecule and a DNA polymerase. The resulting double stranded DNA, where the encoded molecule is displayed at the end, is shown in the figure of the example.

35 *Example 4 (see example 4, figure 4):*

In this example of the reaction in stage 2 of the invention, the carriers employed carry double-stranded oligonucleotide identifiers (generated for example by ligation of double-stranded oligonucleotide identifiers during the Stage 1 synthesis). Thus, these carriers are brought together and reacted without the use of a template. The oligonucleotides of the

carrier are constructed so that the resulting double stranded DNA of one repertoire of carriers (e.g. X) has a "sticky" end (a 3' or 5' overhang), which is complementary to a corresponding "sticky" end of another repertoire of carriers (e.g. Y). Any two carrier molecules from each repertoire can therefore be combined by annealing using the sticky ends of the double stranded DNA oligos. The carriers are hereafter ligated and reacted like in the other examples. The identifier DNA sequence of the combined molecules will, in addition to the 48-mer coding sequence, contain a "sticky" end derived sequence that will be identical for all molecules. One advantage of this method is that synthesis of long templates is avoided.

10 After reaction of the molecule fragments the single stranded DNA can be made double stranded by a standard extension reaction using a DNA polymerase and a primer annealing to the end of the DNA.

In examples 1-4 it may be desirable to include primer binding sites, either at the ends of the template DNA strand, or as part of the identifier (one primer binding site per carrier) when stage 2 involves a ligation of two carriers to each other, to form a complementary template. These primer binding sites can be used for amplification of the template (or complementary strand), allowing generation of more copies of the encoded molecules.

20 *Example 5 (see example 5, figure 4):*

Carrier molecules with double-stranded oligonucleotide identifiers, as provided in example 4, can also be ligated into a circular double stranded DNA molecule, by digestion of the double stranded DNA molecule with restriction enzymes, which creates "sticky ends", and ligation to corresponding "sticky ends" on the double-stranded oligonucleotide identifiers.

25 The double stranded DNA molecule used for ligation of the oligonucleotide identifiers can be a plasmid having a replication origin, which enables it to be transformed into cells, e.g. bacteria, for amplification. The double stranded DNA molecule can also be a non-functional piece of DNA only used to make circular DNA molecules containing the double-stranded oligonucleotide identifiers.

30 In the first step the carrier molecules from one repertoire of synthesis in stage 1 are ligated into the circular DNA. In the second step, the carrier molecules from another repertoire of synthesis in stage 1 are ligated into the double-stranded DNA molecule, adjacent to where the carrier molecules from the first step were inserted. The carrier molecules are hereafter reacted as in the other examples.

35

Synthesis of the template:

10¹² different templates are required in order to generate a library of 10¹² encoded molecules. In the example shown in figure 1 and 2, the library consists of tetrameric

polypeptides that are encoded by a 48-mer template sequence. Each of the four amino acid units in an encoded molecule is encoded by a 12-mer sequence-unit of the DNA template. In the examples, the number of different amino acids that are encoded by the 12-mer sequence-units is 1000. A simple and efficient way of synthesising the 10¹² DNA

5 templates in the example is thus by synthesising the 1000 different 12-mer oligonucleotides, corresponding to the first position of the template. This set of position 1 oligonucleotides are then incubated with a corresponding set of position 2, 3 and 4 oligonucleotides, and ligated, to form templates comprising four 12-nucleotide sequence units. A practical approach to this is suggested in figure 5:

10 i) Four sets of double-stranded DNA fragments are generated, with overhangs at both ends. The overhangs may be of any length, for example one-nucleotide overhangs. Thus, for each of the four sets, each of 1000 different 13-nt DNA oligo sequences are incubated with a partially complementary 13-nt DNA oligo. 12 of 13 nucleotides are complementary, and therefore duplex DNA with one-nucleotide overhangs at both ends is generated.

15 ii) Then the 4 sets each of 1000 duplex DNA fragments are incubated, and ligated together by a DNA ligase to form 10¹² different DNA templates, each comprising four sequence-units of 12 base pairs. A set of primer annealing sequences can subsequently be ligated to the ends of the templates, in order to obtain templates that can easily be amplified by PCR using primers complementary to the flanking sequences. Alternatively, as shown in figure 20 5, the primer-annealing-sites may be carried by the terminal sequence units (in this example comprising position 1 and position 4 sequence-units). The primer binding sites also allow sequencing of the templates.

iii) Single-stranded template is generated by removing the upper or lower strand, for example by including biotinylated position 1 upper strand oligos. Streptavidin beads may be added, which will immobilise the double stranded templates on the beads. Then alkaline conditions are employed to melt the DNA duplexes, which releases the lower strand from solid support. The single stranded DNA templates may now be used in the Stage 2 DNA-templated synthesis reactions. Alternatively, the DNA-templated synthesis may be performed using the immobilised upper strand as template, and thus performing the reactions on solid support. This should allow the DNA-templated reactions to be performed under aqueous conditions as well as in organic solvents.

The library of DNA templates may be generated in other ways. For example, 10¹² different templates may be generated by standard oligonucleotide synthesis of fully random oligonucleotides, or oligonucleotides with constant, partly randomised, and/or fully randomised positions. Alternatively, the template may be generated by split and mix synthesis, as described below.

The nucleic acid component of bi-functional molecules generated in stage 1 and stage 2 synthesis.

- The present invention involves oligonucleotide identifiers of three different uses. Thus, 5 "unit identifier" is the smallest unit, and typically describes the part of the final encoded molecule that is added and becomes attached to the nascent bi-functional molecule as the result of a synthesis round during stage 1 synthesis. "Carrier identifier" describe the oligonucleotide that anneals to the template during a stage 2 synthesis round. Finally, "template identifier" or "identifier template" describes the encoding oligonucleotide of the 10 bi-functional molecule. Thus, in the example described in figures 1 and 2, there are 4 unit identifiers that are linked two and two to generate 2 carrier identifiers. Then, the two carrier identifiers are hybridised to 1 identifier template carrying four identifiers complementary to the identifier portions of the carriers. In this example, the encoded molecule is attached to the identifier template non-covalently by annealing of the 15 oligonucleotide that carries the encoded molecule to the identifier template.
- In most of the examples, two molecule fragments and two unit identifiers are used to generate a carrier molecule, but the number of molecule fragments and unit identifiers used to synthesise a carrier can also be 1, 3, 4, 5, 6, 7 or higher. The resulting molecules 20 are termed carrier molecules, where a given carrier molecule contains molecule fragments that are linked together, and where the linked molecule fragments are also linked to the identifiers that have also been linked together.
- The number of carriers that bind to the same template may be 1, 2, 3, 4, 5, 6, 7 or more.

The identifiers of carrier molecules must be capable of hybridising to the template in a 25 sequence- or partly sequence-specific way. Thus, in a preferred embodiment the template is a nucleic acid or nucleic acid analogue, and the identifiers of the carrier molecules consist of DNA, RNA, PNA, LNA, or other oligonucleotide analogues capable of sequence specific hybridisation through base pairing. The resulting structure may be a double or triple helix.

30 The templates employed during stage 2 synthesis likewise must be capable of hybridising to the carriers in a sequence-specific or partly sequence-specific way. Thus, the templates preferably consist of DNA, RNA, PNA, LNA, or other oligonucleotide analogues capable of sequence specific hybridisation through base pairing. In a preferred embodiment, the template is amplifiable (i.e. can be amplified through the use of a polymerase, such as is used in a PCR-reaction, or by repeated replication such as in a cell). Example amplifiable templates are DNA and RNA, and unnatural DNA- or RNA, capable of being used as a template in a polymerase-based transcription or replication process.

The carrier identifiers preferably anneal sequence specifically to the templates. Therefore, the carrier identifiers should be of a length and composition that allows a relatively strong

and specific interaction between template and carrier under the hybridisation step.

Preferably, the length of the carrier identifier is in the range of 3-50 nucleotides, more preferably 7-25 nucleotides, and most preferably 8-20 nucleotides. In order to ensure specific annealing of the carriers, i.e. to ensure that the correct carrier anneal to a given 5 template, the set of carrier identifiers should be chosen such that the overlap between identifiers of different carriers is as insignificant as possible. In other words, the design of the set of carrier identifiers should ensure a high degree of non-identity among any carrier identifier in the set. Likewise, the templates should be designed so that interaction with undesired carrier identifiers is minimised.

- 10 Preferably, less than 30% of the nucleotides of a given carrier identifier sequence should be complementary to a non-desired hybridisation sequence of the template. Less preferably, less than 50%, and even less preferably, less than 70%, and the least preferably, less than 90%, of the nucleotides of a given carrier identifier sequence should be complementary to a non-desired hybridisation sequence of the template. This will 15 ensure a high degree of specificity during the annealing step in the templated synthesis, and hence, ensure that the desired carriers bind to the template, and hence, ensure that the desired encoded molecules are generated. In principle, even where two carrier identifiers are different only at one nucleotide position, the stage 1 and stage 2 synthesis and encoding/decoding should work. However, the more different any two identifiers in a library are, the more robust the encoding, selection and characterisation will be.

In addition, the more different the identifier template sequences are, the less sensitive the system will be to errors introduced during the chemical reactions or PCR amplifications performed. For example, if all unit identifier pairs have different nucleotides at more than 5 positions, two nucleotide substitutions introduced during PCR will still allow the correct 25 identification of the unit identifier. Therefore, the non-identity preferences that are described above for carrier identifiers also applies to unit identifiers employed during stage 1 synthesis, also if no stage 2 synthesis is performed to generate the bi-functional molecules employed in the screening step.

- 30 In order to obtain efficient attachment of the incoming identifier and the nascent bi-functional molecule during stage 1 synthesis, it is preferable to use oligonucleotide identifiers that have constant regions, i.e. a portion of the identifier is the same for all of the identifiers in the library. In the stage 1 synthesis, this region could be relatively long (e.g., 20 nucleotides), to ensure efficient annealing to a complementary oligonucleotide, and hence, efficient ligation. However, if the bi-functional molecule produced in this way is to be used as a carrier molecule in a subsequent stage 2 synthesis, the constant regions should be as short as possible, in order to ensure high annealing specificity among the different carrier molecules being used in the stage 2 synthesis. An appropriate compromise 35 between these two opposing factors is to include constant regions (for mediating the

ligation of unit identifiers) that are preferably between 1 and 10 nucleotides long, less preferably 1-20 and least preferably 1-50 nucleotides long. Most preferably, the constant region is 2, 3, 4, 5, or 6 nucleotides long. In order to ensure efficient annealing and thus ligation, but to reduce the length of the constant region, the constant region may include 5 unnatural oligonucleotides such as e.g. LNA, which provides a higher affinity than natural DNA.

Variations and specifications to the general scheme described above for the generation of bi-functional molecules. A number of methods by which to generate carriers exist. In

10 principle, any number of methods exist by which to perform the templated reactions. Almost any combination of methods for carrier generation and templated synthesis can be applied to the generation of bi-functional molecules. Below specific approaches to carrier generation have been outlined (Sub-procedures 1-X) and specific approaches to templated synthesis have been outlined (Sub-processes A-Z). Individual processes 1-X and A-Z, as

15 well as combinations 1A, 1B, ..., 1Z, 2A, 2B, ..., 2X, 2B, ..., 2X, 2B, ..., 2X can be applied to the generation of bi-functional molecules.

Specific enablements of stage 1 synthesis, or generation of carriers, include but is not limited to:

20

This is a variation of bi-functional molecule formation described in (Lerner et al., EP 0643778 B1, Encoded combinatorial chemical libraries), but any embodiment of bi-functional molecule synthesis described in said patent is applicable for stage 1 synthesis.

Two alternating parallel combinatorial syntheses are performed so that a genetic tag is 25 chemically linked to a polypeptide or other type of organic molecule being synthesised; in each case, the addition of one amino acid residue (or other type of molecule fragment with at least one reactive group) to the structure is followed by the addition of an oligonucleotide sequence, which is defined to code for that amino acid (or molecule

30 fragment), i.e., to function as an identifier for the structure of the amino acid residue (or molecule fragment). The library is built up by the repetition of this process after pooling and division. During the process, therefore, a bi-functional molecule is being formed, consisting of a polypeptide (or other organic molecule) attached to an identifier tag that encodes the synthetic history of the encoded molecule, and hence, encodes the expected chemical structure of the encoded molecule. Thus, the process is an example of a split-35 and-mix DNA tagging process.

In a related process, also described in (Lerner et al., EP 0643778 B1, Encoded combinatorial chemical libraries), a method is described for preparing tagged molecules or libraries of tagged molecules, where each molecule is attached to a solid support. The method comprises a solid support, being dispersible in aqueous solution, a first linkage

unit coupled to the solid support, a second linkage unit coupled to the first linkage unit, a bi-functional unit coupled to the second linkage unit, wherein the bi-functional unit (or "linker") has two reactive groups, one of which is employable for oligonucleotide synthesis (i.e., may be used as an initiator-functionality for oligonucleotide synthesis), and one of 5 which is employable for polypeptide- (or other organic molecule-) synthesis.

This is a variation of bi-functional molecule formation described in (Dower et al., EP 0604552 B1), but any embodiment of bi-functional molecule synthesis described in said patent is applicable for stage 1 synthesis. Organic molecules are synthesized in a 10 component by component fashion (i.e. by a split-and-mix-like process) on solid support or particles. During the synthesis of the organic molecule, a tag is simultaneously synthesised which becomes linked to the organic molecule via a linker. The organic molecule-oligonucleotide tag complex may be released from these supports to provide a soluble library of bi-functional molecules.

15

This is a variation of bi-functional molecule formation described in (Freskgård et al., WO 2004/039825 A2), but any embodiment of bi-functional molecule synthesis described in said patent application is applicable for stage 1 synthesis. A bi-functional carrier molecule is generated as follows: A nascent bi-functional complex comprising a reactive group and a 20 priming site for enzymatic addition of an oligonucleotide identifier is reacted at the reactive group with one or more molecule fragments, and provided with respective identifier(s) identifying the molecule fragments at the priming site using one or more enzymes.. The enzyme used may be a DNA ligase, and the priming site may be a 3'-OH group on a

nucleic acid to which an oligonucleotide tag carrying a 5'-phosphate can be attached. In a 25 preferred embodiment, the ligation of identifiers does not involve an enzyme. Enzymes, including the DNA ligase, are in general substrate specific, entailing that the enzymatic addition of a tag to the priming site is not likely to interfere with the display molecule being formed. Generally, the bi-functional carrier molecule is formed by more than a single round of reaction between one or more molecule fragments and the reactive group. In a certain

30 aspect of the invention, the nascent bi-functional complex reacted with one or more molecule fragment(s) and provided with respective oligonucleotide identifier(s) is reacted

further one or more times with one or more molecule fragment(s) and is provided with respective identifier(s) to produce an encoded molecule as one part of the bi-functional carrier molecule and an encoding part comprising identifiers which codes for the identity of 35 the molecule fragments which have participated in the formation of the encoded molecule.

The reaction at the reactive group and the addition of identifiers may occur in any order, i.e. the reaction may occur subsequent to, simultaneously with, or previous to the identifier addition. The choice of order may among other things be dependent on the enzyme type, the reaction conditions, and the type of reactant. The encoding part of the nascent bi-

functional complex is formed by addition of at least one unit identifier to a priming site using one or more enzymes. Further unit identifiers may be attached to a previous unit identifier so as to produce a linear or branched encoding part. As long as at least one unit identifier is attached by an enzymatic catalysed reaction, further unit identifiers may be provided using chemical means or enzymatic means at the discretion of the experimenter.

The identifier can be added to the priming site using any appropriate enzyme. In a certain embodiment, an identifier is provided at the priming site of the nascent bi-functional complex utilizing an enzymatic extension reaction. The extension reaction may be performed by a polymerase or a ligase or a combination thereof. The extension using a

10 polymerase is suitably conducted using a complementary oligonucleotide that hybridises to the previous unit identifier, and carries a sequence complementary to the unit identifier; the unit identifier is then synthesized by e.g. a polymerase by using the complementary oligonucleotide as template, and the end of the previous unit identifier as starting point for the extension. Typically, the previous unit identifier would carry a 3'-OH at the end, and a

15 polymerase would extend from this 3'-OH, using dNTPs or NTPs. Examples of suitable enzymes for the addition of unit identifiers to the nascent bi-functional molecule include DNA polymerase, RNA polymerase, Reverse Transcriptase, DNA ligase, RNA ligase, Taq

DNA polymerase, Pfu polymerase, Vent polymerase, HIV-1 Reverse Transcriptase, Klenow fragment, or any other enzyme that will catalyze the incorporation of complementing

20 elements such as mono-, di- or polynucleotides. Polymerases that allow mismatch extension can also be used, such as for example DNA polymerase α (Washington et al., (2001) JBC 276: 2263-2266), DNA polymerase β (Vaisman et al., (2001) JBC 276: 30615-

30622), or any other enzyme that allow extension of mismatched annealed base pairs. In another aspect, when ligases are used, suitable examples include Taq DNA ligase, T4 DNA

25 ligase, T4 RNA ligase, T7 DNA ligase, and E coli DNA ligase. The choice of the ligase depends to a certain degree on the design of the ends to be joined together. Thus, if the ends are blunt, T4 RNA ligase may be preferred, while a Taq DNA ligase may be preferred for a sticky end ligation.

30 This is another variation of bi-functional molecule formation described in (Freskgård et al., WO 2004/039825 A2). In this approach, molecule fragments are attached to the unit

identifier or an oligonucleotide that is complementary to the unit identifier. Using this approach, the synthesis of a library of many different bi-functional carrier molecules can be conducted in a single vessel. In contrast to the split-and-mix synthesis, where the reaction

35 and unit identifier addition must be carried out in separate compartments. Thus, this variation entails a method comprising the steps of i) providing a nascent bi-functional complex comprising a reactive group and an oligonucleotide identifier region, ii) providing a molecule fragment-oligonucleotide conjugate comprising an oligonucleotide sufficient complementary to the oligonucleotide identifier to allow for hybridisation, a transferable

molecule fragment, and a complementary anti-codon identifying the functional entity, iii) mixing the nascent bi-functional molecule and the molecule fragment-oligonucleotide conjugate under hybridisation conditions to form a hybridisation product, iv) transferring the molecule fragment of said conjugate to the nascent bi-functional molecule through a

5 reaction involving the reactive group of the nascent bi-functional complex, and v) enzymatically extending the oligonucleotide identifier to obtain a unit identifier attached to the bi-functional molecule having received the molecule fragment. The enzymatic extension may occur subsequent to or simultaneously with the transfer of the functional entity or even prior to the transfer.

10

This is a variation of bi-functional molecule formation described in (Lerner et al., EP 0643778 B1, "Encoded combinatorial chemical libraries" and Dower et al., EP 0604552 B1). In this method, a stage 1 synthesis is performed in which identifiers are ligated together in the absence of an enzyme, but with the aid of a nucleic acid that brings the

15 identifiers into close proximity, and hence increases ligation efficiency this way (see Figure 12). An identifier that must be ligated must comprise, e.g. at one end of the identifier

oligonucleotide, a chemical group that can react with a chemical group on the identifier that it will be ligated to. As an example, the nucleic acid may be complementary to a part of the sequence of both identifiers. In a preferred embodiment a nucleic acid is added that

20 is complementary to the ends of the identifiers. Hybridisation of this bridging nucleic acid to the two identifiers brings the two chemical groups, one from each identifier, into close proximity, thereby improving the ligation reaction efficiency. or nucleic acid analog, or any other type of molecule or solid support that brings the two identifiers into proximity and thereby increases the efficiency of ligation.

25 The complementary nucleic acid brings the ends of the two identifiers into close proximity; the ends of the identifiers have been modified to make them prone to reaction under these conditions. As an example of such chemical ligation, one of the identifiers may carry a 3'-

30 hydroxyl group, and the other identifier carry a 2-methylimidazole-activated phosphate at its 5'-end. When brought into proximity by hybridisation to a complementary nucleic acid, the 3'-OH and the 5' activated phosphate will react to generate a native phosphodiester bond between the identifiers. This and other types of chemical ligation reactions that can be used to ligate identifiers together chemically are described in more detail above.

Practically, any two reactive groups that may react to form a bond between the two identifiers can be used, as long as the ligation conditions do not modify the identifiers to an extent that abolishes efficient hybridisation of the carriers in the stage 2 synthesis that follows carrier synthesis. Thus, as an example, the reactive groups and the reactions listed

35 in figure 6 and 7 may be used for ligation of identifiers as well.

The encoded molecule formed using this approach, and the reactive groups and types of reactions that may be used to generate the encoded molecule, are the same as those

mentioned for general stage 1 synthesis (and stage 2 synthesis), and thus include reactions and reactive groups listed in figure 6, 7 and 8.

This variation of combinatorial chemistry combines the combinatorial chemical synthesis of 5 molecule fragments with a tagging of each molecule synthesized. Combinatorial chemistry

is employed to synthesize a library of molecule fragments, whereafter the molecule fragments are attached to oligonucleotides, to produce bi-functional molecules, that may be used as carriers in templated synthesis. In this approach, combinatorial chemistry is applied to the generation of a library of molecule fragments. Any type of combinatorial

10 chemistry may be used, for example split-and-mix synthesis on beads (see review by

Abelson (1996) *Methods in Enzymology*, vol. 267, p. 211-221; Lebl et al. (2000) *US*

6,090,912; Lebl et al. (1998) *US* 5,840,485), followed by the release from the bead to

produce multiple solutions (e.g. in separate wells), each containing a specific molecule

fragment, or the library may be generated in an array format, for example on glass or pins

15 (Dapremont et al. (1995) *Physiol. Chem. Phys. & Med. NMR*, 27: 339-343, "Multiple

synthesis using the multipin method"), before it is eventually released from solid support,

to produce multiple solutions (e.g. in separate wells), where each well contain a specific

molecule fragment. Any other means of generating a library of compounds can be applied,

including the embodiments of (Still et al. (1998) *US* 5,721,099; Dower et al. (1991) *US*

20 1991000762522; Boger et al. (2001) *US* 6,194,612 B1; Cook et al. (2001) *US* 6,191,273;

Gustafson et al. (2000) *US* 6,140,361; Graybill et al. (2000) *US* 6,127,191; Dervan et al.

(2000) *US* 6,090,947; Baidur et al. (1999) *US* 5,891,737; Baidur et al. (1997) *US*

5,646,285). Once the library of molecule fragments is generated, each specific molecule

fragment is attached a specific oligonucleotide. As a result, a library of bi-functional

25 molecules have been generated, where specific molecule fragments are attached to

specific oligonucleotides that thus encode the structure of the molecule fragment

generated by combinatorial chemistry. The bi-functional molecules may be used as carriers

in a templated synthesis.

In a preferred embodiment, the oligonucleotides are prepared by combinatorial synthesis,

30 either before or after their addition to the solutions containing the specific molecule

fragments. Thus, for the tagging of n^2 molecule fragments, n oligonucleotides of length 12

nt are pairwise attached, for example by chemical ligation, to n oligonucleotides of length

13 nt, to produce n^2 oligonucleotides of length 25 nt that is then attached to the n_2

molecule fragments, for example by an acylation reaction involving an amino group on one

35 of the ends of the oligonucleotides, and a carboxylic acid of the molecule fragment. For

example, if a library of 10^6 different molecule fragments have been prepared, one may

generate 10^6 different oligonucleotides by pairwise ligation of two sets of 10^3

oligonucleotides.

This is a variation of the principle described immediately above, where combinatorial chemical synthesis is replaced by any synthetic method that produces a library of more than one hundred, one thousand, or ten thousand molecule fragments. In this variation it

is thus not a requirement that a combinatorial approach is employed for the synthesis of

5 the library of molecule fragments. Again, the bi-functional molecules are generated by

attachment of specific oligonucleotides to each of the molecule fragments, and again, the

oligonucleotides can be attached to the molecule fragments directly, or after a

combinatorial ligation of a smaller set of oligonucleotides has produced a larger set of

oligonucleotides that may then be attached to the molecule fragments. Again, the

10 combinatorial ligation of oligonucleotides can also be performed after attachment of the

first oligonucleotide to the molecule fragment.

This is a variation of any of the other stage 1 synthesis principles, or formation of carrier molecules (1-8), in which identifiers are linked by a non-covalent bond. As an example, the

15 identifiers may contain complementary oligonucleotide sequences. In a preferred

embodiment, a strong, non-covalent association of the identifiers by extended

complementarity may allow hybridisation of the hybridised identifiers (the carrier

molecule) to a template in a subsequent templated synthesis.

20 This is a variation of Morgan et al., 2005, WO 2005/058479. Any embodiment of bi-

functional molecule synthesis described in said patent application is applicable for stage 1

synthesis. The invention provides a method of synthesising libraries of molecules which

include an encoding oligonucleotide tag. The method utilises a "split and pool" strategy in

which a solution comprising an initiator (similar to the linker molecules described in this

25 patent application), comprising a first building block linked to an encoding oligonucleotide,

is divided into multiple fractions. In each fraction, the initiator is reacted with a second,

unique, building block and a second, unique oligonucleotide which identifies the second

building block (the oligonucleotides are reacted through the use of an enzyme, e.g. a

30 ligase). These reactions can be simultaneous or sequential and, if sequential, either

reaction can precede the other. The dimeric molecules produced in each of the fractions

are combined and then divided again into multiple fractions. Each of these fractions is then

reacted with a third unique building block and a third unique oligonucleotide which encodes

the building block (the oligonucleotides are reacted through the use of an enzyme). The

number of unique molecules present in the product library is a function of the number of

35 different building blocks used at each step of the synthesis, and the number of times the

pooling and dividing process is repeated.

This is a variation of Harbury and Halpin (WO 00/23458). Any embodiment of bi-functional molecule synthesis described in said patent application is applicable for the synthesis of bi-

functional carrier molecules. The method involves the synthesis of a plurality of compounds, comprising a) forming a first group of subsets of identifier templates, where the templates in each subset each has a selected one of a plurality of different first hybridisation sequences, a mixture of different second hybridisation sequences, and a reactive group, b) reacting the reactive group in each of the subsets formed in (a) with a selected molecule fragment, thereby to form a molecule fragment-specific compound intermediate on the associated sequence in each subset, c) forming a second group of subsets of the reacted templates, where the templates in each subset each have a selected one of a plurality of different second hybridisation sequences, and a mixture of different first hybridisation sequences; and d) reacting the compound intermediates in the sequences in each of the subsets formed in (c) with a selected molecule fragment.

The method may be performed on solid support, for example by sorting the templates according to the sequence of e.g. the first hybridisation sequence by annealing to beads, each of which carry a specific oligonucleotide that is complementary to a specific first hybridisation sequence, to allow the formation of different subsets of templates, where each subset of templates have a specific hybridisation sequence. The method thus provides subsets of nucleic acid templates, generated by base-specific duplex formation between each different first hybridisation sequence and a complementary oligonucleotide. The reactive group in each of the subsets are reacted with a selected molecule fragment to form a molecule fragment-specific compound intermediate (i.e., a bi-functional molecule comprising a specific molecule fragment). The end result thus is, after grouping the templates into subsets each comprising a particular first hybridisation sequence, and reacting each subset with a specific molecule fragment, and repeating this process for the second, third, fourth, etc. hybridisation sequence, a number of bi-functional molecules, comprising an encoded molecule (comprising n linked molecule fragments) and an encoding template identifier, comprising n hybridisation sequences (here called unit identifiers).

The templates employed can carry 1, 2, 3, 4, 5, 6, 7 or more hybridisation sequences, and thus, the bi-functional molecules generated by this approach can carry 1, 2, 3, 4, 5, 6, 7 or more molecule fragments. The number of different molecule fragments in a given round can be from 2 to 10,000. The template may be amplifiable or non-amplifiable by polymerases such as Taq polymerase.

The method, as described above, as well as the embodiments described in (Harbury and Halpin, WO 00/23458) may be performed using the chemistries and protecting groups described in this patent application, including chemistries and protecting groups shown in table 6, 7 and 8, and using any type, composition and length of oligonucleotides described in this patent application.

Specific enablements of stage 2 synthesis (templated synthesis of bi-functional molecules) include but is not limited to:

This is a variation of the principle described in (Bruck et al., Chemistry and Biology, January 1996, 3: 49-56, "Template-directed ligation of peptides to oligonucleotides"). In

5 this approach a number of identifier templates are incubated with two (or more) sets of carrier molecules. One set of carriers comprise a molecule fragment comprising a reactive group A (for example an activated ester of a carboxylic acid, for example a thioester), and one set of carriers comprise a reactive group B (e.g. the amine). Thus, the templated a primary or secondary amine, capable of reacting with the thioester). Upon

10 template/carrier complex formation, an acylation reaction takes place, leading to the transfer of the molecule fragment comprising the reactive group A (e.g. thioester) onto the molecule fragment comprising the reactive group B (e.g. the amine). Thus, the templated reaction leads to formation of an encoded molecule consisting of two molecule fragments, linked by an amide bond.

15 Any type of reactive groups may be used in this method, including reactive groups mentioned in this patent application, including table 6, 7, and 8. For amine acylations, thioesters, N-hydroxysuccinimide esters, and phenol esters are particularly well suited for the direct transfer of one molecule fragment onto another. Alternatively, an indirect acylation between a carboxylic acid and an amine may be mediated by EDC, EDC/NHS, DMT/MM and other reagents that activate the acid for nucleophilic attack. After covalent linkage of the two molecule fragments, one of the linkers is cleaved, to transfer one molecule fragment onto the other molecule fragment, associated with the same template. In addition, many other reactions may be used to link two molecule fragments in an oligonucleotide-templated fashion. Thus, this principle may be applied to the generation of

25 encoded molecules as listed in the present invention, and may involve reactive groups mentioned in the present invention, including those listed in figure 6, 7, and 8.

This is a variation of the principle described in (Walder et al. (1979) Proc. Natl. Acad. Sci. USA, 76: p. 51-55), and also is closely related to the principle described immediately

30 above. In this approach, the formation of an amide bond between two molecule fragments is facilitated by the juxtaposition of these by hybridisation of two oligonucleotides, each carrying one molecule fragment, to a complementary template. An oligonucleotide, in the present invention termed a carrier identifier, to which a carboxylic acid of a molecule fragment is attached through ester linkage, is hybridised to a template, which is also

35 hybridised to a second oligonucleotide carrying a molecule fragment comprising an amino group. The duplex DNA is designed so as to bring the ester and the amine into close proximity. Because of this proximity, a reaction between the ester and the amine takes place; this leads to formation of an amide bond between the two molecule fragments, and the ester is cleaved to allow transfer of the ester-bound molecule fragment onto the

molecule fragment comprising the amino group. An ester linkage is thus used to allow direct transfer of the molecule fragment; in principle any substitution-labile acyl linkage should allow the direct transfer of one molecule fragment onto the other in this scheme. In addition, many other reactions may be used to link two molecule fragments in an

5 oligonucleotide-templated fashion. Thus, this principle may be applied to the generation of encoded molecules as listed in the present invention, and may involve reactive groups mentioned in the present invention, including those listed in figure 6, 7, and 8.

This is a variation of (Liu et al. (2002), WO 02/074929 A2, "Evolving new molecular function"). Any embodiment of bi-functional molecule synthesis described in said patent application is applicable for the stage 2 synthesis of bi-functional molecules. In a preferred embodiment of this application, the method comprises first providing one or more nucleic acid templates, which one or more nucleic acid templates optionally have a reactive unit associated therewith. The nucleic acid template is then contacted with one or more

15 transfer units designed to have a first moiety, an anticodon, which hybridises to a sequence of the nucleic acid, and is associated with a second moiety, a reactive unit, which includes a building block of the compound to be synthesised. Once these transfer units have hybridised to the nucleic acid template in a sequence-specific manner, the synthesis of the chemical compound can take place due to the interaction of reactive moieties

20 present on the transfer units and/or the nucleic acid template. Significantly, the sequence of the nucleic acid can later be determined to decode the synthetic history of the attached compound and thereby its structure. The method allows the synthesis of large numbers of molecules using combinatorial methods. The principle may be applied to the generation of encoded molecules as listed in the present invention, and may involve reactive groups mentioned in the present invention, including those listed in figure 6, 7, and 8.

D. This is a variation of the templated synthesis described in (Pedersen et al. (2002) WO 02/103008 A2, "Templated molecules and methods for using such molecules"). Any

30 embodiment of bi-functional molecule synthesis described in said patent application is applicable for the stage 2 synthesis of bi-functional molecules. This variation involves the generation of a library of bi-functional molecules, where the individual bi-functional molecule synthesis comprises the steps of i) providing at least one identifier template comprising a sequence of n coding elements, complementary to n carrier identifiers, ii) providing a plurality of carriers, wherein each carrier comprises a) at least one carrier identifier oligonucleotide

35 capable of recognising a predetermined coding element, b) at least one molecule fragment with at least one one reactive group, and c) at least one linker separating the at least one molecule fragment from the at least one carrier identifier, iii) contacting each of said coding elements with a carrier identifier capable of recognising said coding element, and iv) obtaining a bi-functional molecule comprising covalently or non-covalently linked molecule

fragments by linking, by means of a reaction involving reactive groups, two or more molecule fragments, wherein the bi-functional molecule is linked by means of a linker to the identifier template or the complementary template that templated the synthesis of the bi-functional molecule.

5 The principle may be applied to the generation of encoded molecules as listed in the present invention, and may involve reactive groups mentioned in the present invention, including those listed in figure 6, 7, and 8.

E. This is a variation of the principle described in (Pedersen et al. (2003) WO03/078625 A2, "An improved method for synthesizing templated molecules"). Any embodiment of bi-functional molecule synthesis described in said patent application is applicable for the stage 2 synthesis of bi-functional molecules. The variation provides a method for synthesising a bi-functional molecule, said method comprising the steps of: a) providing at least one identifier template comprising one or more codons (i.e., oligonucleotide sequences

15 complementary to carrier identifiers), b) providing a first carrier molecule comprising a zipping domain, said zipping domain comprises a first part of a molecule pair, being capable of reversible interaction with a second part of the molecule pair, c) providing one or more second carriers, comprising a zipping domain comprising the second part of said molecule pair and capable of reacting with the first carrier, d) contacting the components of step a), b), and c) with each other under conditions allowing specific hybridisation of the carriers to the template(s) and dimerization of the two parts of the molecule pair, e) allowing the reactive groups of the first carrier to react with the reactive groups of the second carrier(s), f) optionally, cleaving one or more linkers, provided that at least one linker remains to connect the molecule fragments (i.e., the encoded molecule) with the template or the complementary template, g) obtaining a bi-

20 functional molecule where the identifier template directed the synthesis of the encoded molecule.

The principle may be applied to the generation of encoded molecules as listed in the present invention, and may involve reactive groups mentioned in the present invention, including those listed in figure 6, 7, and 8.

30 F. This is another variation of the principle described in (Liu et al. WO 2004/016767 A2, "Evolving New Molecular Function"). Here, the templated synthesis is performed with a template that has an "omega" or "O" type architecture. This type of template permits distance-dependent nucleic acid-templated reactions to be encoded by bases far removed from the associated reactive unit. The method involves providing (i) a template comprising a first reactive unit associated with a first oligonucleotide comprising a codon and (ii) a transfer unit comprising a second reactive unit associated with a second oligonucleotide comprising an anti-codon that is capable of annealing to the codon. The codon and/or the anti-codon include first and second regions spaced apart from one another. The oligonucleotides then are annealed together to bring the reactive units into reactive proximity. When the oligonucleotides anneal to one another, the codon

(or anti-codon) with the spaced-apart regions produce a loop of oligonucleotides not annealed to the corresponding anti-codon (or codon). A covalent bond-forming reaction then is induced between the reactive units to produce the reaction product, the "encoded molecule".

5

G. These approaches are alternative methods for the templated stage 2 synthesis, and are described in the figures 2-8, 11 and 13, and are described in detail elsewhere in this application.

10

H. This is a variation of the approach for stage 2 synthesis, in which the template and the carrier molecules do not have to be associated (hybridized) during the reaction of the molecule fragments (Franch et al., WO 2004/083427, "Ligational encoding of small molecules"). In the course of the encoding process, a single-stranded product is generated, including both the template and some or all of the carrier molecules. Thus, the molecule fragments, and the reactive groups that react during the stage 2 synthesis step, are covalently associated with the template, which allows e.g. a higher temperature to be employed during the reaction step. This may be an advantage for reactions that are particularly enthalpy-driven.

20

Bi-functional molecule formation (i.e. formation of carriers for use in a templated synthesis or formation of the final encoded molecule). The above methods for carrier synthesis and templated synthesis can be combined in any way. As an example, 3A (subprocess 3: formation of bi-functional carrier molecules, as described by any of the embodiments of (Freskgaard et al., WO 2004/039825 A2), for example through enzymatic ligation of identifiers; followed by direct transfer acylation stage 2 synthesis as described in (Bruick et al., Chemistry and Biology, January 1996, 3: 49-56), can be applied to the generation of bi-functional molecules.

25

30 In a preferred embodiment, a one-to-one relationship exists between the identifiers and molecule fragments. Thus, one specific identifier identifies (encodes) one specific molecule fragment. However, it is not essential for the invention that such one-to-one relationship exists. For example, during stage 1 synthesis different identifiers may be added to the same well, wherefore more than one identifier will encode one molecule fragment. It is still possible, however, to identify the molecule fragment directly from the sequence of the identifier. Likewise, it is possible to add several molecule fragments to the same well during stage 1 synthesis. In this case, however, the identity of the molecule fragment cannot be deduced directly from the sequence of the attached identifier, but must be deduced by other means, for example mass spectrometry. Finally, it is also possible to add

the same identifier to several wells, or the same molecule fragment to several wells. In the former case, however, it is not possible to identify the molecule directly from the sequence of the identifier.

5 Screening methods employing bi-functional molecules.

Once the bi-functional molecules have been generated, the desired molecules may be identified in any way possible. Thus, a number of screening methods exist, for the identification of organic molecules with desired characteristics. Different types of selection or screening protocols are described in (Liu et al. (2002), WO 02/074929 A2; Pedersen et al. (2002) WO 02/103008 A2; Pedersen et al. (2003) WO03/078625 A2; Lerner et al., EP 0643778 B1, Encoded combinatorial chemical libraries; Dower et al., EP 0604552 B1; Freskgaard et al., WO 2004/039825 A2; Morgan et al., 2005, WO 2005/058479; Harbury and Halpin, WO 00/23458).

10

15 Specific screening methods employing bifunctional molecules for the identification of organic molecules with desired characteristics include but are not limited to:
Affinity selection on immobilised target molecules. In this approach the target molecules (e.g., DNA, RNA, protein, peptide, carbohydrate, organic or inorganic molecule, supramolecular structure or any other molecule, is immobilized covalently or non-

20

covalently to a solid support such as beads, the bottom of a well of a microtiter plate, a reagent tube, a chromatographic column, or any other type of solid support. A library of bi-functional molecules are now incubated with the immobilized target molecule, excess non-bound bi-functional molecules are washed off by the replacing supernatant or column buffer with buffer not containing bi-functional molecules one or more times. After washing the bound bi-functional molecules are released from solid support by addition of reagents, specific ligands or the like that results in the elution of the bi-functional molecule, or the pH is increased or decreased to release the bound bi-functional molecules, or the identifier of the bi-functional molecule is cleaved off from the encoded molecule with a reagent, pH change or light-induced cleavage. The recovered identifiers can now optionally be amplified by PCR, and cloned and sequenced to reveal the structure of the ligands encoded by the identifier, or alternatively, be amplified and taken through an additional round of templated synthesis. As an alternative, the identifiers or bi-functional molecules comprising identifiers, are not released from solid support, but rather the identifiers are optionally amplified by PCR directly while still immobilised on solid support.

30

35 Affinity selection on target molecules in solution, followed by any means of isolation of the bi-functional molecules bound to the target, e.g. by immunoprecipitation of the target-bi-

functional molecule complexes. A library of bi-functional molecules are incubated with target molecules (e.g. a protein). After complex formation of bi-functional molecules with target, the complex is isolated from non-complexes, for example by the addition of

35

polyvalent antibodies against the target molecule and precipitation of antibody-target-bi-functional molecule complexes, or is precipitated by the addition of beads that bind the target molecules. The latter may for example be by addition of streptavidin-coated beads that bind to pre-biotinylated targets. The identifiers recovered by precipitation can now be characterized or amplified, e.g., by PCR, as described in (I). The sequence of the identifiers will reveal the identity of the encoded molecules that bind the target molecules.

Affinity selection on target molecules in solution, followed by gel retardation, chromatographic separation e.g. size exclusion chromatography, or separation by centrifugation e.g. in a CsCl_2 -gradient. A library of bi-functional molecules are incubated with target molecules (e.g. a protein). After complex formation of bi-functional molecules with target, the complex is isolated from non-complexes, for example by gel

electrophoresis or size exclusion chromatography, or any other chromatographic or non-chromatographic method that separates the target-bi-functional molecule complexes from non-complexed bi-functional molecules, for example based on the difference in size and/or charge. The identifiers of the bi-functional molecules of the column fraction or band on the gel that comprises target-bi-functional molecule complexes are now characterised or amplified, e.g., by PCR, as described above. The sequence of the identifiers will reveal the identity of the encoded molecules that bind the target molecules.

Affinity selection on surfaces. Particles, preferably small particles, of solid material, e.g., metal particles, metal oxide particles, grinded plastic, wood, preformed carbon nanotubes, clay, glass, silica, bacterial biofilm or biofilm of other microorganism, cement, solid paint particles, laminate, stone, marble, quartz, textile, paper, skin, hair, cell membranes, industrial membranes, epiderm, or the like, is added to a solution comprising a library of bi-functional molecules. After incubation, one or more washing steps are performed, to

remove unbound bi-functional molecules. Then, the bi-functional molecules bound to the surface, or the identifiers of the bi-functional molecules bound to the surface, are released as described above, and the identifiers characterised and/or amplified as described above.

Selection for intracellularisation. Bi-functional molecules are incubated with cells or micelles, or on one side of a lipid membrane, or on one side of a cell monolayer (e.g.

CaCo2 cell monolayer), in order to allow the bi-functional molecule to pass or become immobilized into the membranes. Then, a number of washing steps are performed in order to remove bi-functional molecules that have not become immobilized or have passed the membrane. Identifiers from bi-functional molecules that have become immobilized or have passed the membrane are now amplified and/or characterised as described above.

The encoded molecule of bi-functional molecules that have either become immobilized in the membrane or have passed the membrane, represent potential transporters for intracellularization, i.e. by attaching these encoded molecules (without the oligonucleotide tag) to e.g. non-oral drugs these may become orally available, because the transporter mediate their transport across the cell.

Selection by phase partitioning. A two- or three phase system may be set up, wherein the bi-functional molecules will partition out according (at least in part) to the characteristics of the encoded molecules. Therefore, the principle allows the identification of encoded molecules that have particular preference for a certain kind of solvent. Again, the identifiers of the isolated bi-functional molecules can be amplified and/or characterised after the selection has occurred. It may be necessary to coat the nucleic acid component of the bi-functional molecule with e.g. DNA binding proteins, in order to ensure that the partitioning of the bi-functional molecule is significantly correlated with the characteristics of the encoded molecule of the bi-functional molecule.

10 Selection for induced dimerisation of target molecules. In a preferred embodiment, encoded molecules are sought that induce the dimerization of target molecules. For example, small molecules with the potential to induce dimerization of protein receptors in the cell membrane may be applicable as therapeutics. Thus, a selection protocol for encoded molecules with the potential to induce dimerization of proteins A and B is as follows: A library of bi-functional molecules are incubated with proteins A and B. After incubation, the solution is applied to gel electrophoresis, ultracentrifugation (e.g. CsCl -centrifugation), size exclusion chromatography, or any other kind of separation that separates the protein A-protein B-bi-functional molecule-complex from un-complexed protein A and B, and other undesired complexes, such as protein A-protein B-complex. Bi-functional molecules from the band or fraction corresponding to the size and/or charge of the protein A-protein B-bi-functional molecule-complex is recovered, and template identifiers are then amplified and/or characterised as described above. In this case, the encoded molecule would be resynthesized, and tested in a protein dimerisation assay for its effect on the dimerisation of protein A and B.

25 Selection by iterative rounds of binding and elution. This is a modification of the methods reported previously (Doyon et al. (2003), J. Am. Chem. Soc., 125, 12372-12373). Bi-functional molecules are incubated with e.g. immobilised target molecule, e.g. a biotinylated enzyme immobilised on streptavidin beads. After washing one or more times, the bound bi-functional molecules are released from solid support by a change in pH, or by addition of an excess of ligand that binds the target molecule (the ligand can be e.g. a small molecule, peptide, DNA aptamer or protein that is known to bind the target molecule). Alternatively, the bi-functional molecules may be released by degradation of the immobilised target (e.g. by nuclease or protease), denaturation of target or induced conformational changes in target structure or the like. The recovered bi-functional molecules are now re-applied to e.g. immobilised target molecule, optionally after removal or degradation of the ligand or reagent used for elution in the previous step. Again, washing is performed, and the bound bi-functional molecules eluted. The process of incubation and binding, washing and elution can be repeated many times, until eventually only bi-functional molecules of high affinity remains. Then the identifiers of the bi-

functional molecules are amplified and/or characterised. Using this kind of iterative binding and elution, enrichment factors higher than 100.000-fold can be obtained.

Targets may be immobilised on columns, on beads (batch selection), on the surface of a well, or target and ligands may interact in solution, followed by immunoprecipitation of the target (leading to immunoprecipitation of ligands bound to target).

Screening in compartments. A library of template identifiers are incubated with lipids, bi-

functional carrier molecules, primers, nucleotides and other components necessary for micelle formation and PCR amplification, transcription, stage 2 templated synthesis, and assay. The lipids are allowed to assemble into micelles; the number of input templates are adjusted so that on average every fifth micelle contains a template after micelle formation.

A PCR reaction is performed, to generate multiple copies of the same template in each micelle. Then transcription is performed, in order to generate many copies of the corresponding single-stranded RNA. The carriers now hybridise to the RNA, and reaction between reactive groups of the carriers take place, to form the same encoded molecule in multiple copies in a given micelle. Finally, the assay is performed (for example, an enzyme assay that examines the ability of the encoded molecule to inhibit enzyme activity).

Positive micelles are picked by hand under microscope, or sorted using e.g. a FACS sorting machine. Then the identifiers contained within the positive micelles are amplified and/or characterised as described above. Encoded molecules identified in this way represent potential inhibitors of the enzyme; after resynthesis of the free encoded molecule, the encoded molecule can be examined for its inhibitory effect on enzyme activity in standard inhibition assays.

Whole organism selection. A library of bi-functional molecules, optionally modified by e.g. coating proteins, is injected into a dead or living animal, for example a mouse. After

incubation for a period of time (e.g. two hours) in the animal, specific tissue or organs are recovered, and the bi-functional molecules associated with specific organs can be characterised, by e.g. PCR amplification and/or sequencing of the corresponding identifiers. As a specific example, a mouse carrying a tumor can be injected with a library of bi-functional molecules. After incubation, the tumor can be isolated from the animal.

The bi-functional molecules associated with the tumor are potential therapeutics or diagnostics for that cancer.

Any other kind of selection or screening which may be performed employing libraries of bi-functional molecules.

The abovementioned target molecules may be any supramolecular structure (e.g. nanoclusters, multiprotein complex, ribosomes), macromolecule (e.g. DNA, RNA, protein, polymers such as carbohydrates, thiophenes, fibrin), or low molecular weight compound (e.g. cAMP, small peptide hormones, chelates, morphine, drug).

After having performed any of the selections above, the identifiers of the output bi-functional molecules can be amplified, and taken through one more round of stage 2 synthesis. Then, the same or another selection protocol can be performed. This process can be repeated until an appropriately small number of different bi-functional molecules are recovered.

Any combination of stage 1 and stage 2 synthesis may be employed in the generation of bi-functional molecules. Moreover, any screening method may be combined with any combination of stage 1 and stage 2 synthesis schemes. Thus, referring to the numbering above for stage 1 synthesis procedures (1, 2, 3, ..., X) and stage 2 synthesis procedures (A, B, ..., Z) and screening methods (i, ii, ..., n), the following combinations may be employed during the generation and use of bi-functional molecules:

1Ai, 1Aii, ..., 1An, 1Bi, 1Bii, ..., 1Bn, ..., 2Ai, ..., XZn

Furthermore, any sequence of encoding and screening schemes may be applied:

1Ai+1Ai, 1Ai+1Aii, ..., XAn+XAn, 1Bi+1Bi, ..., XZn+XZn (all of which represent two rounds of library generation and screening), and

1Ai+1Ai+1Ai, 1Ai+1Ai+1Aii, ..., XZn+XZn+XZn (all of which represent three rounds of library generation and screening), and

any sequence of, and any number of, library generation and screening rounds.

Finally, in each of the stage 1, stage 2 and screening steps, one or more approaches may be employed. For example, carriers may be generated during stage 1 synthesis, by any combination of approaches 1-X. The carriers may likewise be employed in any combination of stage 2 synthesis approaches. Finally, the bi-functional molecules generated may be screened by any combination of screening approaches i-n. Again, any sequence of encoding and screening schemes may be applied.

In a preferred embodiment, the stage 1 carrier synthesis is only performed once, whereas the stage 2 synthesis is performed any number of times. In this case, each of the stage 2 syntheses may employ the carriers synthesised in the beginning. Thus, a preferred sequence of stage 1 library synthesis, stage 2 library synthesis and library screening is as follows:

5Ai+Ai, where first

carrier synthesis is performed according to (5) above, where the identifiers are ligated together chemically, followed by templated synthesis as described in (A) above, and using the carriers generated immediately above, followed by

screening of the library generated immediately above, by affinity selection on the target molecule as described in (i) above, followed by PCR amplification of the DNA part of the bi-functional molecules recovered from this affinity selection, followed by

templated synthesis as described in (A) above, using the PCR product generated immediately above as DNA templates, and using the carriers synthesised in the first part of the process, and finally followed by

5 screening of the library generated immediately above, by affinity selection on the target molecule as described in (i) above,

to identify bi-functional molecules with affinity for the target molecule.

10 Finally, the encoded molecules with the desired characteristics (affinity for the target molecule) may be identified, by PCR amplification and sequencing of the DNA templates of the recovered bi-functional molecules.

In another preferred embodiment, three rounds of library synthesis is performed, where

15 the first library is generated solely by stage 1 synthesis, the second library is made through a combination of carrier synthesis and templated synthesis, and the third library is made through templated synthesis using carriers synthesised in the previous round of library synthesis. One such sequence of library syntheses and screening is 5iv+5Aiv+Alv:

First a library of bi-functional molecules is made where four molecule fragments are linked together to form the encoded molecule, and four identifiers are chemically ligated together to form the DNA template, as described in (5) above, followed by affinity selection on a surface (for example a suspension of metal-oxide particles in aqueous buffer) as described

in (iv) above, followed by

Carrier synthesis as described in (5) above, where two molecule fragments are linked together and two identifiers are ligated together, to form carrier molecules consisting of

two molecule fragments and two identifiers, followed by

Templated synthesis, where first the DNA templates of the recovered bi-functional molecules from the screening (iv) above are PCR-amplified, and the templates used in a templated synthesis using the carriers generated immediately above. The choice of

30 identifiers and molecule fragments must ensure that an encoded molecule recovered from the first screening round is amplified and turned into a number of copies of the same encoded molecule, attached to a DNA template, even though different methods are used in the first and second round of library generation (i.e., 5 and 5A, respectively). Thus, the final bi-functional molecules generated either initially or after the amplification of

35 recovered templates, contain identical encoded molecules, but may carry different DNA templates. Library generation is followed by

Affinity selection on a surface (in this example, a suspension of metal-oxide particles in aqueous buffer) as described in (iv) above, followed by

Templated synthesis, where first the DNA templates of the recovered bi-functional molecules from the screening (iv) immediately above are PCR-amplified, and the templates used in a templated synthesis using the carriers generated above, followed by

Affinity selection on a surface (in this example, a suspension of metal-oxide particles in aqueous buffer) as described in (iv) above.

5 Finally, the encoded molecules with the desired characteristic (affinity for the surface) may be identified, by PCR amplification and sequencing of the DNA templates of the recovered bi-functional molecules.

10 Reactive groups may be protected and de-protected at various steps during the synthesis of the encoded molecules, in order to ensure that the reactions proceed in the desired order, and links desired molecule fragments in the desired way. Also, before, during or after linkage of the molecule fragments, individual parts of the encoded molecule may be made cyclic, by reaction of reactive groups of different molecule fragments, or by reaction

15 of reactive groups within a molecule fragment.

After or during stage 1 or stage 2 synthesis, the encoded molecule may be modified, for example by hydrogenation, reduction/oxidation, or by reaction with any chemical moiety. If the modification is done on the pool of bi-functional molecules, the modification is

typically not encoded (i.e., no identifiers are added). As for the stage 1 and stage 2

20 syntheses, it is important that the conditions and reagents of the chemical reaction on the encoded molecule does not modify the template or complementary template to such an extent that the template or complementary template cannot be amplified and/or sequenced. Examples of reactions and reagents that may be applied for the modification of the encoded molecule are shown in figure 6 and 7.

25

The split and mix process (Stage 1) that leads to the formation of bi-functional carrier molecules may involve any number of rounds. Preferably the number of rounds is between 1 and 10. Likewise, the DNA-templated synthesis (Stage 2) may involve templates, each of which that hybridise to any number of carrier molecules. Preferably, a template can

30 hybridise to 1-10 carrier molecules simultaneously. Any template, however, has a given maximum of carriers that it can bind. The carriers may be annealed to the template and reacted two at a time, three at a time, four at a time, more than four at a time, or all at once, or one carrier may be reacted with a reactive group on the template (i.e. the carriers are annealed and reacted one at a time, when the nascent encoded molecule is linked to the template). Any combination of hybridisations and reactions may be used until the reactive groups have reacted and the encoded molecule formed.

35 During stage 1 synthesis, a repertoire of molecule fragments can contain different kinds of reactive groups that participate in different coupling reactions. As an example, the first molecule fragment reacted with the functionality on the linker L may be an amino acid.

After coupling to the linker, an amine NH₂ functionality may be available for reaction with the incoming molecule fragment. Thus, the incoming molecule fragment must contain a reactive group capable of reacting with the amine. Example reactive groups of the incoming molecule fragment are carboxylic acid COOH (reacts through an acylation reaction with the amine by addition of e.g. EDC/NHS), aldehyde CHO (reductive amination reaction with amine), sulfonate (alkylation reaction with amine), sulfonyl chloride (sulphonamide formation with amine), and substituted aromates (nucleophilic aromatic substitution). During stage 1 synthesis each molecule fragment is added to a specific well, and therefore, the reaction conditions that are ideal for a particular reaction may be employed in that specific well.

Likewise, during stage 2 synthesis, carriers carrying a variety of different molecule fragments can be employed. For example, if a library of templates all bind one carrier containing an amine reactive group, the other carrier can contain for example a COOH, CHO, sulfonyl chloride or a substituted aromate. The reactions of stage 2 synthesis may not be separated in separate wells, and therefore, the molecule fragments must be compatible with the reaction conditions and reagents/catalysts used for all reactions performed. The reactions can be performed simultaneously if the reactions can be performed under the same conditions, or the reactions may be performed sequentially if the reactions require different conditions and reagents/catalysts.

For both stage 1 and stage 2 synthesis, the number of reactive groups X on the linker, or the number of reactive groups on the nascent bi-functional molecule can vary. If the linker has one reactive group, the molecule fragment that becomes attached to the linker must have at least one reactive group in order to attach further molecule fragments to the nascent bi-functional molecules. If all molecule fragments contain two reactive groups, one reactive group can react with the nascent bi-functional molecule, and one may react with the next molecule fragment that is attached. The resulting structure is a linear structure, like beads on a string. Alternatively, if the nascent bi-functional molecule carries three or more reactive groups, one reactive group may be used for attachment to the nascent bi-functional molecule, and the remaining reactive groups may be attached to different molecule fragments (in the same or in different synthesis rounds), and the resulting structure is branched. Molecule fragments containing more than two reactive groups are termed scaffolds; the structure generated (where at least two encoded molecule fragments in addition to the nascent bi-functional molecule become attached directly to the same molecule fragment) is termed a scaffolded or branched molecule. Typical scaffolds include aromatic structures, benzodiazepines, hydantoins, piperazines, indoles, furans, thiazoles, sterols, diketopiperazines, morpholines, tropanes, coumarines, quinolines, pyrroles, oxazoles, amino acid precursors, cyclic or aromatic ring structures, and many others, all of which must contain at least three reactive groups in order to be considered scaffolds.

If two or more reactive groups of a molecule fragment react directly with the nascent bi-functional molecule, a cyclic structure results.

The molecule fragment repertoires of different rounds of stage 1 synthesis may be of different size, i.e. a different number of molecule fragments may be employed in different rounds. The identifiers added in different wells or different rounds of stage 1 synthesis may be of different length and/or composition. In a preferred embodiment one well contains one specific identifier and one specific molecule fragment. However, in another preferred embodiment more than one molecule fragment is added to each well. This will result in relaxation of the one-template/one-encoded molecule relationship that exist between template sequence and encoded molecule, and can therefore only be pursued to a limited extent. Likewise, in another preferred embodiment more than one identifier sequence is added to each well. If the same identifier is added to different wells containing the same molecule fragment, this will also result in relaxation of the one-template/one-encoded molecule relationship.

Under certain circumstances, the templated synthesis (Stage 2) may be repeated using the bi-functional molecules generated in the templating process as carriers in a second or third templating process. As an example, a templated synthesis reaction is employed that involves ligation of the two carrier molecules, to form a bi-functional molecule in which all the sequence-units identifying the molecule fragments are covalently linked, and where the encoded molecule remains attached to the DNA portion (for an example, see figure 4, example 2). Then the bi-functional molecule resulting from the templated reaction (in figure 4, example 2, the DNA template attached to the encoded molecule XY), may be used in a next templated reaction, using a template that carries two sequence units, one of which is complementary to the DNA portion of the bi-functional molecule generated in the first templated reaction.

Under certain circumstances, it may be desirable to i) generate the bi-functional molecules used in the screening process through split and mix synthesis only, e.g. exclude the stage 2 templating process in the initial bi-functional molecule generation. If desired, recovered bi-functional molecules from the screen may then be amplified using a combination of split and mix synthesis (stage 1) and templated synthesis (stage 2)(see for example figure 2), using the template generated during the bi-functional molecule synthesis described in i) above.

In i) above, a template is generated using a split and mix procedure, involving also the synthesis of an encoded molecule. However, a simpler version of this split and mix approach can be applied to the generation of a library of DNA templates. As an example, to make a library of n³ DNA templates where each DNA template carries 3 codons, and

where each of the 3 positions can contain any of n different codons, perform the following steps: i) Add an aliquot of a mixture of n position1 codon DNA duplexes to each of n wells, ii) To each well, add a specific position2 codon duplex DNA, iii) Ligate the position 1 codon DNA duplex and the position2 codon DNA duplex (f.ex. by the use of a DNA ligase, or an activated phosphate (e.g. Imidazole-phosphatexxxxx, see ref), iv) Mix the contents of all wells, v) Add an aliquot of the mixture of ligated position1-position2 DNA duplexes to each of n wells, vi) To each of the n wells, add a specific position3 codon DNA duplex, vii) Ligate the position3 duplex to the position1position2 duplexes, viii) Mix the contents of all wells.

10 The resulting pool contains n^3 different DNA templates, carrying n different codons at each of 3 codon positions. The DNA duplexes carrying position1 or position3 codons may also carry primer binding sites, to allow amplification by PCR of the generated templates.

The nucleic acid carrying the codons may be RNA, DNA or any other type of nucleic acid or nucleic acid derivative, and it could be single- or double-stranded.

15

It may be desirable, after the synthesis of the bi-functional molecules but before the bi-functional molecules are used in a selection or screening experiment, to make the template double-stranded. This may be done by annealing a primer to the template strand (or its complementary strand), and extend for example using sequenase. This will form a double stranded oligonucleotide template, which is expected to be more inert in the selection experiment than the corresponding single-stranded oligonucleotide, because of the single-stranded oligonucleotide's ability to bind the target as an aptamer.

In a preferred embodiment, the carrier molecules are immobilised during their synthesis.

25 For example, the following types of solid support may be employed for the immobilisation: Sepharose beads with functionalities like-SH, -COOH or -NH₂ groups (where the identifiers or templates are covalently coupled by e.g. disulfide formation, acylation, and acylation, respectively); tentagel beads with functional groups e.g. -SH, -COOH or -NH₂ groups; streptavidin coated beads (where the carriers or templates are covalently coupled to

30 biotin, which in turn can be non-covalently bound to streptavidin); and many other types of supports and functionalities (e.g. polystyrene, polypropylene, agarose (e.g. Hispanagar); diol functionalities, ester functionalities; amide functionalities, glyoxal functionalities).

Advantages of immobilisation of carriers or templates include the easy isolation of single-stranded carriers or single-stranded templates (for use in stage 2 synthesis), easy removal of excess reagents (molecule fragments, identifiers, catalysts, reactants), the ability to conveniently change solvent (for example, change to organic solvent prior to the reaction of the molecule fragments), removal of protection groups, etc.

Finally, the molecule fragments may be immobilised prior to reaction with the growing carrier molecule. This allows for the purification of growing carriers that have reacted with

the immobilised molecule fragments. This in turn allows purification of full-length carriers (i.e., carriers that contain the desired molecule fragments), which will lead to a more efficient stage 2 synthesis.

The carriers and/or templates may be released from the solid support by hydrolysis (e.g. high pH), proteolysis (of a peptide linker), thiol mediated cleavage of disulfide bond, nuclease-mediated cleavage, etc.

Types of encoded molecules:

Different kinds of molecules may be generated and attached to the DNA template that encodes it. Molecules that may be generated by the present invention include small

compact molecules, linear structures, polymers, polypeptides, poly-ureas, polycarbamates, scaffold structures, cyclic structures, natural compound derivatives, alpha-, beta-, gamma-, and omega-peptides, mono-, di- and tri-substituted peptides, L- and D-form peptides, cyclohexane- and cyclopentane-backbone modified beta-peptides, vinyllogous polypeptides,

15 glycopolypeptides, polyamides, vinyllogous sulfonamide peptide, Polysulfonamide conjugated peptide (i.e., having prosthetic groups), Polyesters, Polysaccharides, polycarbamates, polycarbonates, polyureas, poly-peptidylphosphonates, Azatides, peptoids (oligo N-substituted glycinides), Polyethers, ethoxyformacetal oligomers, poly-thioethers, polyethylene, glycols (PEG), polyethylenes, polydisulfides, polyarylene sulfides,

20 Polynucleotides, PNAs, LNAs, Morpholinos, oligo pyrrolinone, polyoximes, Polyimines, Polyethyleneimine, Polyacetates, Polystyrenes, Polyacetylene, Polyvinyl, Lipids, Phospholipids, Glycolipids, polycycles, (aliphatic), polycycles (aromatic), polyheterocycles, Proteoglycan, Polysioxanes, Polysocyanides, Polyisocyanates, polymethacrylates, Monofunctional,

Difunctional, Trifunctional and Oligofunctional open-chain hydrocarbons.

25 Monofunctional, Difunctional, Trifunctional and Oligofunctional Nonaromat Carbocycles, Monocyclic, Bicyclic, Tricyclic and Polycyclic Hydrocarbons, Bridged Polycyclic Hydrocarbones, Monofunctional, Difunctional, Trifunctional and Oligofunctional Nonaromatic, Heterocycles, Monocyclic, Bicyclic, Tricyclic and Polycyclic Heterocycles, bridged Polycyclic Heterocycles, Monofunctional, Difunctional, Trifunctional and Oligofunctional Aromatic Carbocycles.

30 Monocyclic, Bicyclic, Tricyclic and Polycyclic Aromatic Carbocycles

Monofunctional, Difunctional, Trifunctional and Oligofunctional Aromatic Hetero-cycles.

Monocyclic, Bicyclic, Tricyclic and Polycyclic Heterocycles. Chelates, fullerenes, and

any combination of the above and many others. A non-comprehensive list of example generic and specific structures generated by the present invention is shown in (Figure 13).

35 Each of the molecule fragments linked together during the process may be prepared by any kind of synthetic protocol, including standard organic synthesis, prior to the library generation process. Therefore, any type of chemical moiety can be included in a molecule or library of molecules of the present invention. The linkages (bonds) between the

molecule fragments (the linkages that are generated during the library generation process) must be compatible with the presence of the DNA.

The types of bonds that may be generated must be compatible with the presence of the oligonucleotide. Types of linkages that may be generated in the present invention include

5 amide bonds, carbamates, sulfones, sulfoxides, phosphodiester bonds, carbohydrate bonds, ureas, phosphonates, esters, and many others. A non-comprehensive list of such linkage bonds is shown in (Figures 6 and 7), and is listed within the description below.

Following the encoded synthesis of the molecules (whereby one or more bi-functional

10 molecules have been formed), the encoded molecules may be modified in a non-encoded way. For example, the library of molecules may be hydrogenated, acylated, oxidised or reduced, or protection groups may be removed. It will often be necessary to protect chemical motifs that would otherwise participate in the reactions that attach molecule fragments, either through reaction with molecule fragments, reagents, or other nascent bi-

15 functional molecules. These chemical motifs can then be generated at the end of library synthesis, by deprotection reactions after the encoding- and encoded reactions. In principle, any of the reactions listed in figure 6, 7 or 8 may be used after the encoded synthesis, in order to modify the encoded molecules.

20 Use of the method for the synthesis of specific (one or a few) different molecule species. The described methods for organic molecule synthesis may be applied to a) synthesis of compounds of high stereochemical purity, or b) synthesis of compounds without the cumbersome use of protection groups, or c) as a means to increase the yield of synthesis steps that are usually inefficient, for example because the necessary concentrations cannot

25 be achieved in a standard organic synthesis setting. Examples of such uses of the methods is synthesis of a sequence of saccharides, without the use of protection groups.

Use of the library for screening.

30 The bi-functional molecules generated by the present invention may be used to identify encoded molecules with particular characteristics. For example, the bi-functional molecules may be employed in affinity selection experiments, in which bi-functional molecules capable of binding to proteins, DNA, RNA, surfaces, inorganic or organic molecules, and other molecules and substances may be identified. During the affinity

35 selection experiments bi-functional molecules interacting with these molecules or substances may be isolated, and identified by sequencing the DNA portion of the bi-functional molecules. Alternatively, the bi-functional molecules may be screened for catalytic activity, for the ability to interact with other bi-functional molecules, for the ability to become internalised into a cell, for the ability to interfere with conductance or

fluorescence or any other characteristics of another molecule or substance, or for any other characteristics desired. Finally, the library may be screened for the ability to interact with library members of other types of molecules, for example phage-displayed peptides or proteins, in order to identify bi-functional molecules that interact with peptides or proteins

5 of the phage-display library.

The DNA portion of the bi-functional molecules in these different types of screens and selections allows the easy and highly sensitive identification of the encoded small molecule component responsible for its isolation during the screen or selection. In principle, one species of a bi-functional molecule is enough to allow its identification, by first amplifying

10 the DNA component by PCR, followed by sequencing of the DNA.

In addition, the DNA component allows several rounds of screening and amplification to be performed. Thus, after a screening round, the recovered population of bi-functional molecules may be amplified (by amplification of the DNA template, followed by Stage 1 carrier synthesis (or the original preparation of carriers may be used) and Stage 2

15 templated synthesis using the amplified DNA template for bi-functional molecule synthesis). The amplification generates several copies of each of the recovered bi-functional molecules without the need for identification of the recovered bi-functional molecules. The ability to perform several rounds of screening or selection allows efficient screening of very large libraries of bi-functional (OBS) molecules, involving up to at least

20 10^{16} different bi-functional (OBS) molecules.

A typical selection protocol involves the addition of a population (a library) of bi-functional (OBS) molecules to an affinity column, to which a certain molecular target (e.g., a receptor protein or a DNA fragment) has been immobilised. After washing the column, the binders are eluted. This eluate consists of an enriched population of bi-functional molecules with

25 affinity for the immobilised target molecule. The enriched population may be taken through an amplification round (by first amplifying the template and then using the amplified template in a stage 2 templated synthesis of bi-functional molecules), and then be subjected to yet a selection round, where the conditions optionally may be more stringent.

After a number of such selection-and-amplification rounds, an enriched population of high-affine binders are obtained. A typical selection process is illustrated in figure 9. Other types of selection methods that may be used include immunoprecipitation, FACS sorting, mass spectrometry, cell-surface subtraction, in vivo selection (e.g., injection of bi-functional

30 (OBS) molecules into animals and isolation of bi-functional (OBS) molecules from specific tissues) and gel mobility shift assays.

35

Polyvalent display and other means of increasing the likelihood of identifying encoded molecules with weak characteristics.

Under certain conditions the requirements of an encoded molecule, in order to be isolated during the screening step, are too strong, and few or none of the encoded molecules of a

library are expected to fulfill the requirements. Such requirements may be for example high affinity or high catalytic turn-over.

In those cases it may be desirable to employ a multivalent display mode, i.e., to generate libraries of multivalent encoded molecules (either multiple encoded molecules attached to multiple identifiers, or multiple encoded molecules attached to one identifier). During a selection step in which for example an encoded molecule interacts weakly with a target protein, a multivalent encoded molecule may interact with multiple protein targets through the multiple copies of encoded molecules that it contains, and as a result, may bind with higher affinity because of the avidity effect. Likewise, in a screening or selection step for catalytic efficiency, a multivalent encoded molecule may generate more product in a given time, and may be isolated because of this.

A preferred means of generating libraries of multivalent encoded molecules each containing multiple copies of the same encoded molecule, is as follows (figure 11A). First,

15 a library of templates that may be used in stage 2 synthesis is generated. The library of templates may be generated as described in figure 5, whereafter the libraries are then circularised (for example by ligating the two ends of a template). Alternatively the circular templates can be generated as described in figure 4, example 5. A rolling circle amplification is hereafter performed on the library of templates, leading to the generation

20 of a library of templates, where each template now contains multiple copies of the sequence that may be used in stage 2 synthesis to generate an encoded molecule. Secondly, templated synthesis is performed using the templates generated (using carriers optionally generated by stage 1 synthesis, or generated by any other means), leading to multivalent encoded molecules each containing multiple copies of an encoded molecule.

25 Multivalent encoded molecules containing multiple copies of two different encoded molecules may be generated by ligating together sequences two-and-two before circularisation and rolling circle amplification (i.e., ligate two templates together, circularise the ligation product, and perform rolling circle amplification). After templated synthesis on these templates, the library will consist of bi-functional molecules each with multiple copies

30 of two different encoded molecules.

The multivalent encoded molecules can now be used in various screening or selection processes. For example, the multivalent encoded molecules may be added to an affinity column, to which target protein has been immobilised with an appropriately high density, so that multivalent encoded molecules may interact with several immobilised targets simultaneously. This will lead to the isolation of bi-functional molecules that contain encoded molecules with affinity for the immobilised target protein.

linker molecule contains two reactive groups, the stage 1 synthesis may lead to the formation of two encoded molecules (figure 11B). Hereafter, the divalent carrier molecules can be used in a templated stage 2 synthesis scheme, for generating a library of divalent template encoded molecules (figure 11C). These principles for the formation of

5 divalent encoded molecules may of course be applied to the generation of trivalent and higher valency encoded molecules, by employing linkers during stage 1 carrier synthesis that carry three or more, respectively, reactive groups.

The divalent encoded molecules generated may be used in screening or selection

10 experiments. For example, a library of divalent encoded molecules may be added to beads to which a target molecule has been coupled with appropriately high density, and an affinity selection experiment performed, leading to the isolation of divalent encoded molecules which contain encoded molecules with affinity for the target molecule. Divalent encoded molecules may be particularly advantageous to use when selecting for affinity to a

15 homodimeric target molecule, or any other target that contains two or more identical binding sites. Relevant targets include membrane proteins such as the Epo-receptor, p53, HER2, Insulin Receptor, many interleukins, palindromic DNA- or RNA-sequences, or fibrin. Divalent encoded molecules containing identical encoded molecules are also appropriate for affinity selection on target molecules with one binding site, where the binding site is

20 partly or fully symmetrical, and therefore allows two identical encoded molecules to interact.

A similar principle may be applied to the generation of bi-functional molecules that carry a helper moiety. For example, when searching for an encoded molecule with affinity for a particular nucleic acid sequence, it may be advantageous to generate a bi-functional molecule that contains a nucleic acid sequence that is complementary to the sequence next to the target nucleic acid sequence, and in this way increase the total affinity of the bi-functional molecule for the target nucleic acid (figure 11D). A similar approach may be applied to the isolation of encoded molecules with affinity for any target molecule with two binding sites, or a binding site that can accommodate two binding moieties. Thus, as an example, if a ligand is known for a binding site in a protein, this ligand may be coupled to the bi-functional molecule, in order to guide the encoded molecule to the target protein, and in order to increase the affinity of the bi-functional molecule (carrying the known ligand) for the target protein (figure 11E). A simple way of attaching the known ligand is by hybridisation, i.e. the encoded molecule is linked to the template and the known ligand is linked to an oligonucleotide that is complementary to part of the template (figure 11F)

35 Similar approaches may be used for isolation of encoded molecules with affinity for a target binding site, where the binding site can be occupied by both the encoded molecule and the known ligand simultaneously (figure 11G). Finally, it may be desirable to increase

the overall affinity of the bi-functional molecule for the target by linking a short oligonucleotide that is complementary to the template of the bi-functional molecule to the target. The short oligonucleotide will then function as a helper moiety that increases the affinity of the bi-functional molecule for the target, by hybridisation of the short oligonucleotide to the bi-functional molecule (Figure 11H).

Selections employing such bi-functional molecules to which have been attached a helper moiety may be applied to affinity selection against all kinds of targets, including protein-heterodimers as well as protein-homodimers, and thus molecular targets include HER2, Insulin-receptor, VEGF, EGF, IL-4, IL-2, TNF-alpha, the TATA-box of eukaryotic promoter regions, and many others.

Dynamic combinatorial library of dimers or trimers of encoded molecules.

The bi-functional molecules of a library may be designed in a way that leads to transient complex formation between 2, 3, or more bi-functional complexes during the screening process. This may be desirable, especially in cases where the libraries that have been generated are relatively small, or in cases where it is desirable to screen a large number of combinations of encoded molecules for synergistic effects. In order to generate transient complexes, the bi-functional molecules may be designed so as to comprise half of a transient interaction pair. For example, a short single stranded oligonucleotide region may be included in the design of the identifiers of the bi-functional molecules that result from the present invention; if some of the bi-functional molecules carry a molecular entity "A" and some other bi-functional molecules of the library carry another molecular entity "B" that interacts transiently, i.e. forms a short-lived complex with, "A", then the two sets of bi-functional molecules of the library will form transient dimers of bi-functional molecules. These transient dimers may then be exposed to a screening process, for example affinity selection, where the dimers are then examined for ability to bind to a certain target. As an example, for each of the species of bi-functional molecules, half of the generated bi-functional molecules carry the oligo sequence 3'-ATGC-5' in the proximity of the encoded molecule, and the other half of the generated bi-functional molecules carry the oligo sequence 3'-GCTA-5'. When all the generated bi-functional molecules are incubated at appropriately low temperature, different combinations of dimers will transiently form, and allow for a feature displayed by the combination of the corresponding two encoded molecules to be selected for. This feature could be the binding of the two encoded molecules of the dimer to bind simultaneously to a target molecule. If appropriately designed, trimers may be (transiently) formed, by formation of triplex DNA between three bi-functional molecules. In this way, all the possible dimers (or trimers) of a pool of bi-functional molecules may be screened for the desired feature. See (Figure 14).

Molecular biological methods applicable to bi-functional molecules.

As the present invention involves the templated synthesis of encoded molecules, most in vitro molecular biological techniques may be applied to the DNA-, RNA- or any other oligonucleotide-portion of the template, and as a result of the encoding by this template, indirectly may be applied to the encoded molecule. Examples of such molecular biological techniques applicable to the encoded molecules of this invention are listed in Figure 10.

Characterisation of encoded molecules identified during a screening of a library of bi-functional molecules.

Once the screening of a library of bi-functional molecules have been done, the isolated bi-functional molecules may be identified by cloning of the oligonucleotide portion, and sequencing. The sequencing may be done by any means, including Sanger sequencing, mass spectrometry-based sequencing, single molecule sequencing, or sequencing by hybridisation to oligonucleotide arrays. The characteristics of the encoded molecules thus identified may now be analyzed, either in its free form (after resynthesis by organic chemistry or after generation of the bi-functional molecule followed by cleavage of the linker that connects the encoded molecule and its identifier) or in its oligonucleotide-linked form (as a bi-functional molecule). In order to analyze the bi-functional molecule carrying the specific encoded molecule, individual templates of the bi-functional molecules isolated during the screening may be cloned (by dilution and PCR in separate wells, or by cloning into vectors and propagation in e.g. E. coli), and then amplified by PCR, to produce many copies of the template that encoded the recovered organic molecule. When the encoding process (Stage 2 templated process) is then performed, many identical copies of bi-functional molecules carrying the specific encoded molecule is generated. The characteristics of the specific encoded molecule, when linked to its identifier template, may then be examined. Example assays used for the analysis of the encoded molecules (in their free form or attached to identifiers) include:

Enzyme inhibition assays

Affinity-determination by competition assays and/or ELISA

30 Cell-based receptor binding assays

Cell-based activity assays, based on the interaction of the encoded molecule with molecular targets on the surface of the cells

Biacore-measurements of molecule-ligand or surface-ligand interactions

Affinity and specificity/selectivity determination using arrays of immobilized targets (e.g. array of 100 immobilized phosphatases), onto which the specific encoded molecule (in free- or oligonucleotide-associated form) is added

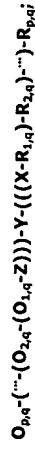
Affinity and specificity/selectivity determination on many specific encoded molecules simultaneously, by immobilization of e.g. 1000 different bi-functional molecules to an array of oligonucleotides, followed by addition of a specific fluorescent protein.

Example assays used for the analysis of the encoded molecules (in their free form) include:

- 5 CaCo2-cell-based analysis of membrane permeability
- In vivo determination of animal toxicity, bioavailability of the compounds, and other ADMET characteristics.
- Solubility of encoded molecule.
- Water-octanol partitioning measurements
- 10 Metabolic stability measurements.

In one embodiment, the reactive groups of a molecule fragment of each of two or more bi-functional molecules hybridized to the same template are reacted. Alternatively, the reactive group of a molecule fragment of one bi-functional molecule is reacted with a reactive group associated with the template to which it is hybridized. Preferably the template is an oligonucleotide template.

In a further embodiment of the invention the number of wells in step a) is m and the number of wells in step f) is n, and may be the same or different for each repetition of steps b) to d) in step f) n; and the structure of the encoded (bifunctional) molecule is



wherein

25 X, Y and Z are components of the linker molecule, L, X being adapted for reaction with a molecule fragment, Z being adapted for reaction with an oligonucleotide and Y being a flexible linker connecting X and Z;

$O_{p,q}$ is the oligonucleotide identifier added in repetition number (p-1) of steps b) to f) in well number q;

30 $R_{p,q}$ is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

p is an integer of at least 1;

m and n are integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and

35 for $O_{1,q}$ and $R_{1,q}$, q is in the range 1 to m, for $O_{p,q}$ and $R_{p,q}$ where p is greater than 1, q is in the range 1 to n.

In another embodiment of the present invention the number of wells in step a) is m and the number of wells in step f) is n, and n may be the same or different for each repetition of steps b) to d) in step f); and the structure of the encoded (bifunctional) molecule is



wherein

$O_{p,q}$ is the oligonucleotide identifier added in repetition number (p-1) of steps b) to f) in well number q;

10 $R_{p,q}$ is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

p is an integer of at least 2;

m and n are integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and

15 for $O_{1,q}$ and $R_{1,q}$, q is in the range 1 to m, for $O_{p,q}$ and $R_{p,q}$ where p is greater than 1, q is in the range 1 to n.

In the formula above the bond to a residue inside a parenthesis () signifies that the bond may be to any part of said residue, e.g. (L- $R_{1,q}$)- $R_{2,q}$ means that $R_{2,q}$ may be bound to either of L and $R_{1,q}$.

20

In an equally important aspect the present invention provides a method for synthesizing an encoded molecule or one or more encoded molecules comprising the steps of:

- a) Dispensing aliquots of a nascent linker molecule L into each of m reaction wells;
- b) Dispensing into each of said m reaction wells a corresponding aliquot of an m^{th} molecule fragment, $R_{1,m}$ and a corresponding aliquot of an m^{th} oligonucleotide, $O_{1,m}$;
- 25 c) Combining all of the nascent bi-functional molecules from all m reaction wells to produce an admixture of nascent bi-functional molecules;
- d) Optionally, Dispensing said admixture of nascent bi-functional molecules into n reaction wells
- 30 e) Optionally, dispensing into each of the n reaction wells of step d) a corresponding aliquot of an m^{th} molecule fragment, $R_{p,q}$, and a corresponding aliquot of an m^{th} oligonucleotide or oligonucleotide identifier, $O_{p,q}$;
- f) Optionally, combining all of the nascent bi-functional molecules from all m reaction wells in step e) to produce an admixture of nascent bi-functional molecules;
- 35 g) Optionally repeating steps d) to f) one or more times;
- h) contacting a resulting bi-functional molecule of step f) or g) with one or more templates, said one or more templates optionally being associated with a reactive

group, under conditions to allow for hybridization of each of the templates to one or more of said nascent bi-functional molecule generated in step f) or g);

- i) Optionally, reacting reactive groups of a molecule fragment of two or more nascent bi-functional molecules hybridized to the same template, or reacting the reactive group of a molecule fragment of one nascent bi-functional molecule with the reactive group associated with the template to which it is hybridized;

the linker molecule L contains at least one reactive group capable of reacting with a reactive group in the molecule fragment and at least one reactive group capable of reacting with a reactive group in the oligonucleotide;

- 10 the molecule fragments each contain at least one reactive group capable of reacting with a reactive group in the linker molecule L or a reactive group in another molecule fragment, and the reactive groups of each molecule fragment may be the same or different;

the oligonucleotide identifiers each contain at least one reactive group capable of reacting with a reactive group in the linker L or a reactive group in another oligonucleotide

- 15 identifier, and the reactive groups of each oligonucleotide identifier may be the same or different;

the oligonucleotide identifier added to each well in step b) and e) identifies the molecule fragment added to the same well in the respective step;

- the steps a) and b) as well as the steps d) and e) may be performed in any order;
- 20 the steps d) and e) in step f) may also be performed in any order.

It is to be understood that the encoded molecule may be a bi-functional molecule with an encoded part comprising one or more molecule fragments, R, and a coding part comprising one or more oligonucleotide identifiers, O.

25

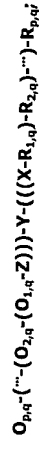
In steps b) and e) of the above process the molecule fragment as well as the oligonucleotide is reacted with the linker to produce a nascent bi-functional molecule.

- In step h) the said bi-functional molecule may be viewed as a carrier or a carrier molecule as mentioned above the template in step i) is preferably an oligonucleotide template.

Again, the process outlined above may be seen as a combination of a step 1 synthesis and a step 2 synthesis, wherein the step 1 synthesis comprises the steps a) to g) and the step 2 synthesis is carried out in step h).

In one embodiment of the method described in the preceding paragraphs the number of wells in step a) is m and the number of wells in step f) is n, and for each repetition of steps b) to d) in step f) n may be the same or different; and

the structure of the encoded (bifunctional) molecule is



wherein

- X, Y and Z are components of the linker molecule, L, X being adapted for reaction with a molecule fragment, Z being adapted for reaction with an oligonucleotide and Y being a flexible linker connecting X and Z;

$O_{p,q}$ is the oligonucleotide identifier added in repetition number (p-1) of steps b) to f) in well number q;

$R_{p,q}$ is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

p is an integer of at least 2;

m and n are integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and

for $O_{1,q}$ and $R_{1,q}$ q is in the range 1 to m, for $O_{p,q}$ and $R_{p,q}$ where p is greater than 1, q is in the range 1 to n.

In another embodiment the method n may for each repetition of steps d) to f) in step g) be the same or different; and
the structure of the encoded (bifunctional) molecule is



wherein

$O_{p,q}$ is the oligonucleotide identifier added in repetition number (p-1) of steps b) to f) in well number q;

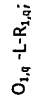
$R_{p,q}$ is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

p is an integer of at least 1;

m and n are integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and

for $O_{1,q}$ and $R_{1,q}$ q is in the range 1 to m, for $O_{p,q}$ and $R_{p,q}$ where p is greater than 1, q is in the range 1 to n. Again, the bond to a residue inside a parenthesis () signifies that the bond may be to any part of said residue, e.g. $(L-R_{1,q})-R_{2,q}$ means that $R_{2,q}$ may be bound to either of L and $R_{1,q}$.

In a further embodiment of the above described method, the structure of the nascent bi-functional molecules resulting from step b) is



and the structure of the nascent bi-functional molecule obtained after repeating the process steps defined in step g) $p-1$ times is

$$O_{p,q} - (') - O_{1,q} - L - R_{1,q} - (') - R_{m,q}$$

wherein p is greater than or equal to 1.9.

5

It is preferred that the number number of reaction wells in step a) and/or the number of reaction wells in step f) is at least 2, such as at least 5, at least 10, at least 25, at least 50, at least 100, at least 200, at least 500, at least 1000 at least, at least 10,000, at least 100,000 at least 1,000,000, at least 10^7 , or at least 10^8 . For certain applications m must be 100,000 or more, such as 10^6 , 10^7 , or 10^8 in step b and c, if the process steps d) to f) are not repeated. If said steps are repeated, however, m may be as low as 10,000 in steps b) and c)

Accordingly it may be preferred to repeat the process steps defined in step g) at least once, such as at least twice, such as at least three times, such as at least four times or 15 more.

In certain embodiments unique identification of the molecule fragments may be preferred.

Accordingly, in these embodiments the oligonucleotide identifier, $O_{p,q}$, added in reaction well number q in repetition number ($p-1$) of the steps specified in step g) uniquely 20 identifies the molecule fragment, $R_{p,q}$, added in reaction well number q in repetition number ($p-1$) of the steps specified in g). In other embodiments, however, identical oligonucleotide identifiers, $O_{p,q}$, are added in two or more reaction wells in the same repetition number ($p-1$) of the steps specified in g).

25 Furthermore, yet other embodiments of the invention involves addition of identical molecule fragments, $R_{p,q}$, in two or more wells in the same repetition number ($p-1$) of the steps specified in g).

Also, it may be preferred to add two or more oligonucleotide identifiers to one or more 30 reaction wells in a repetition of the steps specified in step g). Most preferably, however 1 oligonucleotide identifier is added per well, but it should be recognised that it will be possible to add 2, 3, 4, oligonucleotide identifier per well.

Furthermore the method according to the invention may involve adding two or more

35 molecule fragments to one reaction wells in each repetition of the steps specified in g). Whereas it is preferred to add 1 molecule fragment per well, it will also be possible to

add 2, 3, 4, or 5 molecule fragments per well.

In the method described above, each oligonucleotide identifier comprises a sequence of from 2 to 100, such as from 2 to 90, from 2 to 80, from 2 to 70, from 2 to 60, from 2 to 50, from 2 to 40 from 2 to 30, from 2 to 25, from 2 to 20, from 2 to 15, from 2 to 10 or from 2 to 5 nucleotides, such as at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 nucleotides.

It should be recognised that, preferably, the number of identifiers in each resulting bi-functional molecule is 2. However, each bi-functional molecule may also comprise 1, 3, 4, or 5 identifiers.

10

It should likewise be recognised that each template is preferably capable of hybridising with 2 carriers. Whereas this is preferred, the template may also be capable of hybridising with 1, 3, 4, 5, 6 or 7 carrier molecules.

15 It will for most applications be preferred that the number of nucleotides in each oligonucleotide identifier is 10-15. Other preferred numbers of nucleotides are 5-10, 5-15, 5-20, 10-30.

It should be recognised that the said one or more templates may be an amplifiable

20 template as well as a template, which is non-amplifiable by polymerises. The one or more templates may thus comprise a sequence selected from the group consisting of: nucleotides, unnatural nucleotides, PNA, morpholinos, LNA, RNA, DNA, and other nucleotide analogs capable of base pairing with a natural oligonucleotide or unnatural oligonucleotide. Most preferably the template comprises a DNA sequence or an RNA 25 sequence.

In preferred embodiments of the invention said one or more templates have a length of at least 40 nucleotides, such as 30-50 nucleotides, 20-60 nucleotides, 15 - 80 nucleotides if not considering the primer annealing sites. When these sites are included the total length 30 is preferably approximately 80 nucleotides, such as 70-120 nucleotides, 60-100 nucleotides, 55- 120 nucleotides 50 - 150 nucleotides or 60 - 175 nucleotides. In some embodiments of the invention the said one or more templates may even comprise up to 250 nucleotides or even up to 500 nucleotides.

35 As for the nature of the oligonucleotide sequences each oligonucleotide identifier and/or a sequence of two or more of the oligonucleotide identifiers may comprise a sequence of oligonucleotides, the complementary sequence of which is at least 20% identical, such as at least 30%, such as at least 40%, such as at least 50%, such as at least 55%, such as at least 60%, such as at least 65%, such as at least 70%, such as at least 75%, such as at least 80%, such as at least 85%, such as at least 90% or such as at least 95% identical to the part of said template that hybridises to the identifier. It is to be understood that the

templates when used in the method according to the present invention may be comprised of coding regions separated by spacer regions.

In some preferred embodiments of the present invention the linker, L, is selected from the group consisting of flexible linkers such as Polyethylen glycol, polypeptide, polysaccharide, oligonucleotide, C₆, C₈, C₁₂. In further preferred embodiments of the invention the linker, L, is a cleavable linker. For example the linker may be cleavable by base, acid, light, reagent, heat)

10 In further applications of the method according to the invention the method may involve using the bi-functional product molecule resulting directly from the process steps presented above, that is steps a) to h), as a carrier molecule in a second round of step 2 synthesis. Accordingly, the method may also comprise a further step comprising contacting the bi-functional molecule resulting from step h) with one or more nascent bi-functional

15 molecules and one or more templates each capable of recognizing at least two of the oligonucleotide identifiers present in the bi-functional molecule. The nascent bi-functional are generated through steps a) through f) and the optional steps, g)). It is further to be understood that the identifiers may be covalently or non-covalently linked to each other, and that the identifiers may be double- or single-stranded identifier oligonucleotides, with 20 overhang or blunt-ended.

In a preferred embodiment two or more oligonucleotide identifiers may be covalently linked together, optionally in the presence of a ligase or isomerase enzyme. Alternatively two or more oligonucleotide identifiers are linked by templated extension by enzymes 25 (polymerases).

In some embodiments of the invention two or more oligonucleotide identifiers are ligated by chemical ligation or by a combination of enzymatic and chemical ligation. By a combination of enzymatic and chemical ligation one may understand that two or more oligonucleotide identifiers are ligated by enzymatic ligation in one process step and chemical ligation in a preceding or subsequent process step. For chemical ligation of the oligonucleotide identifiers the following reactions and pairs of reactive groups are the preferred ones:

35 reaction	reactive group	reacts w/	reactive group
phosphodiester formation	3'-OH		Imidazole-activated 5'-phosphate
acylation	amine		ester
acylation	amine		carboxylic acid
disulfide formation	SH		pyridyl-disulfide
40 reductive amination	amine		aldehyde
phosphodiester formation	3'-OH		pyrophosphate-activated 5'-phosphate

tosyl displacement reaction

It is to be understood that the linker of at least one bi-functional molecule may be cleaved simultaneously with or subsequently to hybridisation of the oligonucleotide identifier of 5 said nascent bi-functional molecule to the template.

In the method according to the invention said reactive groups of molecule fragments of the bi-functional molecules or of molecule fragments of the bi-functional molecules and of the template may be reacted in a reaction selected from the group consisting of: acylation, 10 reductive amination, alkylhalide alkylation, Wittig reaction, sulphonylation, isocyanate addition, Suzuki coupling, nucleophilic aromatic substitution, thiourea bond formation, carbamate formation, Heck coupling, HWE reaction, 1,3-dipolar cycloaddition, Michael addition, nitro aldol condensation.

15 In some embodiments of the invention the resulting encoded molecule has a linear structure and is selected from the group consisting of: dimers, trimers, tetramers, pentamers, multimers, and polymers.

In further preferred embodiments at least one molecule fragment having more than one reactive groups has been used in the preparation of the resulting encoded molecule. 20 It is to be understood that the number of reactive groups per molecule fragment may be 1, 2, 3, 4, 5 or 6. It is further to be understood that, in the final product, the reactive groups of said at least one molecule fragment has reacted with other molecule fragments. The resulting molecule is thus a branched or scaffolded structure selected from the group of

25 molecules comprising a scaffold with two substituents, scaffolds with three substituents, scaffolds with four substituents, and scaffolds with five substituents. In this context the term "substituent" means a molecule fragment that has been reacted at the substituent position of the scaffold.

30 In another main aspect the invention provides a method for identifying a molecule with desired characteristics, said method comprising synthesizing a library of encoded molecules by a method as described above.

In some embodiments the method further comprises a step of subjecting the library to a 35 partitioning or enrichment procedure, to identify encoded molecules with desired characteristics. In still other embodiments the method further comprises screening a plurality of encoded molecules in a process to identify and optionally increase the relative amount of an encoded molecule having one or more desired characteristics. Various means

and methods for carrying out enrichment procedures and screening are described in the present specification which may be applied to the process.

In some preferred embodiments the method comprises identifying the encoded molecule 5 by determining the oligonucleotide sequence(s) of the attached identifiers.

In yet another main aspect of the invention a library of encoded molecules or bi-functional molecules is obtained. Said encode or bi-functional molecules may be obtained or may be obtainable by a method according to any of claims.

10

In the library the number of different compounds or compound species may be at least 100, such as at least 1000, at least 10,000, at least 100,000, at least 10⁶, at least 10⁷, at least 10⁸, at least 10⁹, at least 10¹⁰, at least 10¹¹, at least 10¹², at least 10¹³ or at least 10¹⁴.

15

Additional aspects and embodiments of the invention are described in brief in the following:

a) A method for synthesising one or more encoded molecules, comprising the following 20 steps:

step 1: dispensing aliquots of a nascent linker molecule L comprising the components X, Y, and Z, where X is adapted for reaction with a molecule fragment, Z is adapted for reaction with an oligonucleotide and Y is a flexible linker connecting X and Z, into each of m 25 reaction wells; then

step 2: dispensing into each of the m reaction wells of said step 1 a corresponding aliquot of the mth molecule fragment R_{1,m} and a corresponding aliquot of the mth oligonucleotide identifier O_{1,m} to allow reaction between the molecule fragment and X of the linker, and 30 reaction between the oligonucleotide and Z of the linker, to produce a product bi-functional molecule R_{1,m} - L - O_{1,m} where the produced bi-functional molecule comprises a reactive group; then

step 3: combining all of the nascent bi-functional molecules from all m reaction wells 35 produced in said step 2 for producing an admixture of nascent bi-functional molecules; then

step 4: dispensing equal aliquots of the admixture of nascent bi-functional molecules from the prior step into each of m reaction wells, then

step 5: dispensing into each of the m reaction wells of said step 4 a corresponding aliquot of the mth molecule fragment represented by R_{n,m} and a corresponding aliquot of the mth identifier molecule represented by O_{n,m} for producing a nascent bi-functional molecule represented by :

$$5 \quad R_{n,m} - (\dots) - R_{1,m} - L - O_{1,m} - (\dots) - O_{n,m}$$

wherein n is greater than or equal to 2; then

step 6: combining all of the elongated nascent bi-functional molecules from all m reaction wells of step 5 for producing an admixture of elongated nascent bi-functional molecules; 10 then

step 7: repeating steps 4-6 until the desired bi-functional carrier molecules, each formed from the reaction of n molecule fragments and n oligonucleotide identifiers, are produced, and where each bi-functional carrier molecule generated comprises one or more reactive 15 units; then

step 8: providing one or more templates, which one or more templates optionally have a reactive unit associated therewith; then

step 9: contacting one or more carrier molecules of step 7 with said one or more templates under conditions to allow for specific hybridisation of the oligonucleotide identifiers of the one or more carrier molecules to the one or more templates; then

step 10: reacting the reactive units of the molecule fragments of at least two carrier 25 molecules hybridised to the same template, or reacting the reactive unit of a carrier molecule with the reactive unit associated with the template to which it is hybridised, to generate one or more encoded molecules.

30 b) The method as described in a), wherein the oligonucleotide identifiers of two or more carrier molecules are covalently linked together prior to, during or after step 9 or 10.

c) The method as described in b), wherein the oligonucleotide identifiers are covalently linked together after step 9 but before step 10, by a ligase enzyme.

35

d) The method of b), wherein the oligonucleotide identifiers are covalently linked together after step 9 but before step 10, in the absence of a ligase enzyme.

e) The method of a) to d), wherein the template is dissociated from the carrier molecules prior to reaction between said reactive units.

f) The method of a) to e), wherein more than one molecule fragment is added to the same reaction well in steps 2 or 5, to allow for a multiple component reaction to take place.

g) The method of a) to f) where m does not have the same value in different repetitions of step 5.

h) The method of a) to g), where in step 2 or 5 the oligonucleotide identifiers are linked to the nascent bi-functional molecule by a ligase enzyme.

i) The method of a) to h), where in step 2 or 5 the oligonucleotide identifiers are linked to the nascent bi-functional molecule without the use of a ligase enzyme.

j) The method of a) to i), wherein a library of more than one bi-functional molecule is generated, the method further comprising enriching for library members comprising an encoded molecule displaying a desired property.

k) The methods of a) to j) wherein steps 4-7 have been eliminated.

l) The methods of a) to k), wherein steps 8-10 have been eliminated.

m) The library of two or more bi-functional molecules generated by the method of a) to i).

n) The library of 10^6 or more bi-functional molecules generated by the method of a) to i).

o) The library of 10^{10} or more bi-functional molecules generated by the method of a) to i).

With respect to the above description of the various aspects of the present invention and of the specific embodiments of these aspects it should be understood that any feature and characteristic described or mentioned above in connection with one aspect and/or one embodiment of an aspect of the invention also apply by analogy to any or all other aspects and/or embodiments of the invention described.

When an object according to the present invention or one of its features or characteristics is referred to in singular this also refers to the object or its features or characteristics in

plural. As an example, when referring to "a cell" it is to be understood as referring to one or more cells.

Throughout the present specification the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps.

10 FIGURE LEGENDS

Figure 1. Reactions in stage 1 of the method in the invention

Schematic representation of an example of the synthesis steps of stage 1. Two rounds of "split and mix" synthesis are shown leading to the generation of bi-functional carrier molecules each carrying a different di-peptide and a unique 24-mer oligonucleotide that encodes the di-peptide. Each round of synthesis adds an amino acid and an identifier oligonucleotide. (m) represents the number of different molecule fragments in each of the two different repertoires employed. (m) can have a different values for different repertoires.

The split and mix synthesis shown in the example includes the following steps:

Add linker molecule to wells 1-m

Add amino acids $R_{1(1-m)}$ to wells 1-m, and react with linker.

25 Add oligonucleotides $O_{1(1-m)}$ to wells 1-m, and react with the linker.

Mix content of wells 1-m and split into 1-m wells on a new plate.

Add amino acids $R_{2(1-m)}$ to wells 1-m and react with reactive group of $R_{1(1-m)}$

Add oligonucleotides $O_{2(1-m)}$ to wells and react with $O_{1(1-m)}$.

Mix content of wells 1-m

30

Figure 2. Reactions in stage 2 of the method in the invention

Example of the synthesis steps of stage 2. In the illustrated example the bi-functional carrier molecules generated in the example in figure 1 are combined by a template directed method. Tetra-peptide bi-functional carrier molecules with 48-mer identifier oligonucleotides are therefore generated in the example. (m) represents the number of different molecule fragments in each of the four different repertoires, i.e., in the example $m = 1000$ for all four repertoires. (M) represents the total number of (encoded) molecules generated. Here, $M = (1000)^4 = 10^{12}$.

The synthesis in the example comprise the following steps:

Add bi-functional carrier molecule from stage 1

Add DNA templates that bind the carriers through their complementary oligo's

Acyl transfer reaction where the amino group of di-peptide in one carrier attacks the

5 peptidyl ester of the di-peptide in the other carrier

The synthesis is complete

Figure 3. Types of molecule fragment transfer from one carrier to another:

Direct transfer reaction: The reaction between reactive groups leads directly to the transfer
10 of a molecule fragment. The mechanism is shown schematically ("generic") as well as for a specific case ("example"). Other types of reactions allowing direct transfer are shown in figure 6.

Indirect transfer reaction. Reaction between reactive groups leads to the formation of a linkage between the two reactive groups. Thereafter a molecule fragment is cleaved off
15 from one carrier, mediating its transfer to another carrier. "Civ" indicates a cleavable moiety, i.e. a part of the linker that is cleavable, for example by acid, base, electromagnetic radiation, light, heat, or by specific reagents or catalysts.

Long horizontal line symbolises a template. Short horizontal line symbolises oligonucleotide identifier.

20

Figure 4. Alternative methods for the reactions in stage 2

Examples of alternative methods for carrying out the reactions of stage 2 are shown. A number of templates (long horizontal line) is mixed with sets of carriers (short horizontal line). In the examples, two sets of carriers are employed. The first example is identical to
25 the example shown in figure 2, except that in this example the two carriers are ligated together before reaction of the reactive groups of the two carriers. The next two examples show variations of the template directed reactions illustrated in the first example. In

Example 2, the identifiers of the two carriers are ligated together prior to reaction of the reactive groups of X and Y, and the duplex structure denatured, to generate single-stranded complementary template. The single-stranded structure improves the likelihood
30 of X and Y reacting. The reaction efficiency of the reactive groups of X and Y may be increased by including complementary sequences next to X and Y. This will lead to stable duplex formation proximal to X and Y, positioning X and Y in close proximity, and thereby increasing the reaction efficiency. In Example 3, one of the carriers is ligated to the

35 template through its oligonucleotide identifier moiety. The ligation of carrier to template may be stimulated by including a hair-pin structure in the template, as shown in the figure, and then ligating together the template and carrier by use of for example a ligase or chemical ligation. Example 4 shows a template-free method of carrying out the reactions in stage 2. The carriers of the example are double-stranded, allowing for efficient

ligation of their overhangs. Ligation of the carriers lead to the formation of a complementary template. Before reaction of X with Y, the duplex structure is denatured, allowing a more efficient reaction of X and Y. As in example 3, the efficiency of reaction may be increased by including complementary sequences proximal to X and Y,
5 respectively.

The library of encoded molecules that results from each of the examples can be of the same kind; however, the examples describe different set-ups that may allow different chemical reactions to be performed. In a preferred embodiment, the template of the bifunctional molecules are turned into double-stranded DNA before selection or screening
10 is performed, in order to eliminate potential interaction from the single-stranded regions. Thus, in examples 2 and 4, it may be advantageous to add a terminal oligonucleotide that anneals to the DNA template that carries the encoded molecule; by extension, e.g. by a polymerase such as Sequenase, a double-stranded DNA will be generated, carrying the encoded molecule at one end.

15 The lower strand of the duplexes symbolises the template (long horizontal line); short horizontal lines symbolise oligonucleotide identifiers; after ligating the oligonucleotide identifiers together, a complementary template is formed, symbolised by a long horizontal line. X and Y are molecule fragments, each containing at least one reactive group.

20

Figure 5. Template generation:

An example of a DNA library generation process for the synthesis of a library of 10^{12} DNA templates. Four sets of 1000 DNA oligos are mixed individually with their complementary sequences, to form for example 12 nt duplex DNA, with an overhang of one nucleotide at both ends. In the example each oligo carries a region complementary to an identifier
25 sequence (a "codon" sequence). In addition, the oligos of the two distal sets of oligos contain constant regions. These constant regions may be used at a later stage for PCR-amplification, sequencing or polymerase extension. This is followed by a ligation step in which the overhangs mediate the ligation of the 4000 duplex DNA complexes, to form 10^{12} (=1000⁴) different duplex DNA complexes. Ligation may be by a ligase or by chemical

30 ligation. Optionally, the templates may be amplified by e.g. PCR using primers that anneal to the constant regions at the ends of the template, or by any other molecular biological technique that allows amplification. Finally, one of the strands is isolated in order to be used in a templating process (stage 2 templated synthesis). The isolation of single stranded templates can be done in a number of ways, including asymmetric PCR on the

35 ligated product (which leads to excess of one of the strands), or by including a biotinylated PCR-primer that anneals to one end of the template and thus leads to incorporation of biotin into one of the strands of the duplex; by immobilisation of the biotin on streptavidin-coated solid support, and denaturation of the duplex template, one may recover the non-

biotinylated strand from the supernatant, or the biotinylated strand immobilised on solid support.

In the example, the configuration of each of the 10^{12} single-stranded templates thus is as follows: "Constant sequence-codon1-codon2-codon3-codon4-Constant sequence". Each of 5 the codon positions contain one specific of the 1000 possible sequences, i.e., each template carries its specific combination of a codon1-, codon2-, codon3-, and codon4 sequence.

Templates may also be generated by stage 1 split and mix synthesis, in which optionally the reaction of molecule fragments with the carrier is excluded. To generate a DNA 10 template library as the one described in this example, 4 sets of 1000 different duplex DNA molecules must be ligated, employing 1000 wells in each of four rounds of split and mix synthesis. This will generate the same DNA template library of 10^{12} molecules as described above.

15

Figure 6. Direct transfer: Reactive groups and bonds formed upon reaction:

A number of reactions are shown that mediate the direct transfer of molecule fragments from one carrier to another. In the left part of the figure the two carriers (the donor- and the acceptor carrier) are shown. The oligonucleotide identifiers of the carriers are indicated 20 by a short horizontal line. The templates to which the carriers bind are indicated by long horizontal lines. The carriers carry molecule fragments containing reactive groups that upon reaction lead to the transfer of a molecule fragment from one carrier onto the other. The reactions that allow direct transfer include acylation (formation of amide, pyrazolone, isoxalone, pyrimidine, coumarine, quinolon, phthalhydrazide, diketopiperazine, hydantoin, 25 benzodiazepinone, etc), alkylation (including reductive amination not shown in figure), vinylation, disulfide formation, addition to carbon-hetero multiple bonds, such as Wittig/Wittig-Horner-Emmon (formation of substituted alkenes), transition metal catalysed reactions such as arylation (formation of biaryl, vinylarene), alkylation, nucleophilic substitution using activation of nucleophiles, such as condensations, and cycloadditions. All 30 of these reactions may be used for indirect transfer as well. The reactions may also be used during stage 1 synthesis. Figure 6 is adapted from (Pedersen et al. (2002) WO 02/103008 A2, "Templated molecules and methods for using such molecules").

Figure 7. Indirect transfer: Reactive groups and bonds formed upon linking reaction:

Indirect transfer involves first the coupling reaction between the reactive groups of carrier 35 molecules, followed by a cleavage that releases one molecule fragment from its carrier molecule. This figure shows examples of reactive groups that may for example be used in the coupling reaction. The coupling reaction may be nucleophilic substitution, aromatic nucleophilic substitution, transition metal catalysed reactions, addition to carbon-carbon

multiple bonds, cycloaddition to multiple bonds, and addition to carbon-hetero multiple bonds. In figure 8 a number of example cleavable linkers that may be combined with these coupling reactions in order to obtain efficient indirect transfer are shown. The reactions may also be used during stage 1 split and mix synthesis. Figure 7 is adapted from 5 (Pedersen et al. (2002) WO 02/103008 A2, "Templated molecules and methods for using such molecules").

Figure 8. Cleavable linkers and protection groups, cleaving agents and cleavage products:

10 Cleavable linkers and protection groups that may be used to release molecule fragments in an indirect transfer reaction, or can be used as protecting groups, are shown. The cleavable linkers may be combined with reactive groups from figure 7, in order to indirectly transfer molecule fragments from one carrier to another. Linkers may be cleaved by acid, base, electromagnetic radiation, light, heat, by specific reagents or catalysts such as LiOH, 15 PdCl_2 , TCEP, NaIO_4 , etc. Figure 8 is adapted from (Pedersen et al. (2002) WO 02/103008 A2, "Templated molecules and methods for using such molecules").

Figure 9. A typical affinity selection process:

An example affinity selection process is shown. First a DNA template library is generated, for example as described in figure 5. Then, stage 2 templated synthesis is performed using 20 the carriers generated in stage 1 (not shown), which generates a library of bi-functional molecules. The target may be biotinylated, allowing its immobilisation on magnetic beads coated with streptavidin. The beads are immobilised on a magnet and washed. The bound ligands are then eluted, and the DNA of the eluted bi-functional molecules are amplified, 25 for example by PCR, where after this amplified DNA can be used in yet another round of bi-functional molecule library synthesis, or may be sequenced in order to identify the ligand structures that bound to the target.

Figure 10. Molecular biological techniques applicable to bi-functional molecules:

30 A number of molecular biological techniques are listed that allow small molecule engineering, analogous to protein engineering through modification of the DNA encoding the protein. Using bi-functional molecules, one may here modify the encoded small molecule through modifications of the DNA encoding the small molecule. Shuffling of the DNA templates (and hence, the small molecules), can be done efficiently by e.g. restriction endonuclease cleavage of the DNA template in the spacer that separates the codons. Other techniques such as DNA arrays of bi-functional molecules are also suggested. Figure 10 is 35 modified from (Pedersen et al. (2002) WO 02/103008 A2, "Templated molecules and methods for using such molecules").

Figure 11. Polyvalent display and other approaches to the identification of molecules with weak binding characteristics.

Polyvalent display by rolling circle amplification of templates before templated reaction.
 5 DNA template molecules are circularised by ligation of the ends. Specific primers are annealed and extended by rolling circle amplification resulting in templates having multiple copies of the specific binding sites for carrier molecules. The multiple copy templates are thereafter used for templated synthesis with carrier molecules resulting in polyvalent display of encoded molecules.

10

Stage 1 synthesis of divalent bi-functional carrier molecules.

Split and mix synthesis is carried out as in the example describing stage 1 synthesis, but the linker molecule (L) employed in the first step of the synthesis has, in this example, two reactive ends to which molecule fragments ($R_{1..n}$) can be coupled. This results in the
 15 generation of divalent bi-functional carrier molecules having two encoded molecules attached to a linker that is attached to a single oligonucleotide identifier ($O_{1..n}$).

Stage 2 templated synthesis employing divalent bi-functional carrier molecules (generated in example B above).

20 The divalent carrier molecules from example B can be used for templated synthesis employing the method described for stage 2 of the present invention. As a result a library of divalent encoded molecules are generated. Each molecule consists of two encoded molecules ($R_{1..n}$) attached to a linker that is attached to one oligonucleotide identifier ($O_{1..n}$).

25 Template assisted binding to target DNA molecule.

For screening of a library of encoded molecules for binding to target DNA sequences, hybridisation of complementary DNA sequences (C and C') on the bi-functional molecule and the target DNA, can increase the overall affinity and help in the identification of molecules in the library with low affinity for the target DNA.

30

Use of known ligand for assisted target binding.

A library of encoded divalent bi-functional molecules, each containing a known ligand (L) and an encoded molecular entity, (R), is used for screening for molecules with two binding sites, of which one of these is specific for the known ligand. Binding of the known ligand to
 35 its site on the target molecule (T), assists the binding of the encoded molecular entity to the other binding site.

Use of known ligand for assisted target binding – hybridisation of known ligand to template.

This example uses the same principle as illustrated above in figure 11 F, but in this example the known ligand is hybridised to the bi-functional molecule through hybridisation of complementary DNA sequences (C' and C) carried by the known ligand and the template DNA. Hybridisation of the known ligand to the template DNA of the bi-functional molecule
 5 creates functionally divalent molecules that can be used for screening for target with two binding sites, of which one is specific for the known ligand.

Use of known ligand for assisted target binding – binding to the same site.

As in the example illustrated in figure 11 E, but in this example the known ligand (L) and
 10 the molecular entity (R) bind to the same site of the target molecule.

Use of complementary DNA attached to the target molecule to assist binding.

A DNA sequence, C', which is attached to the target molecule, is complementary to a DNA sequence, C, on the template DNA. Hybridisation of the complementary DNA sequences
 15 assists binding of the encoded molecules to the target (T).

20 Figure 12. Example set-ups allowing improved ligation of identifiers.

During the stage 1 synthesis, the identifiers are ligated together. In order to make this an efficient reaction, the identifiers can be double-stranded DNA with overhangs that are complementary, and therefore bring the reactive groups of the identifiers into close proximity. Alternatively, the identifiers are single-stranded and a complementary oligonucleotide, or some other kind of molecule that binds to the identifiers and brings the reactive groups into proximity, is added in order to increase the efficiency of the chemical or enzymatic ligation.

Ligation assisted by "sticky" ends of the DNA

Ligation assisted by complementary oligonucleotide

30 Ligation assisted by complementary oligonucleotide attached to solid support

Ligation assisted by annealing to self-complementary sequence

Ligation assisted by DNA binding molecule

35

Figure 13. Example molecule fragments and the encoded molecules resulting from stage 1 and stage 2 synthesis.

A1-A4 show generic structures of molecule fragments, carrying at least 1 reactive group (A1), two reactive groups (A2), three reactive groups (A3), and four reactive groups (A4). R can be any molecular entity, and can be cyclic or non-cyclic, aliphatic or aromatic. X, Y and A are reactive groups. Molecule fragments can carry multiple reactive groups of the same kind (e.g., three X reactive groups), or can carry multiple reactive groups of different kinds (e.g., X, Y and A).

B1-B4 show specific examples of molecule fragments. B1 structures carry at least one reactive group (here: carboxylic acid or amine). B2 structures carry at least two reactive groups (hydroxyl, amine, thiol). B3 structures carry at least three reactive groups (amine, disulfide, carboxylic acid). B4 carry at least four reactive groups (hydroxyl, amine, thiol, carboxylic acid).

C, D, E, and F show examples of molecules generated by stage 1 and/or stage 2 synthesis, i.e., by covalently coupling molecule fragments through their reactive groups. The stippled circles indicate molecule fragments that have been linked together during the stage 1 and/or stage 2 synthesis.

During stage 1 synthesis the molecule fragments become attached to the linker molecule L via reaction of a reactive group of the molecule fragment with a reactive group of the linker. In this example, the hydroxyl of the encoded molecule of (F) could have been attached to a carboxylic acid-modified oligonucleotide, thus linking the encoded molecule to the linker.

Figure 14. Dynamic combinatorial library of dimers or trimers of encoded molecules.

A library, A, of encoded bi-functional molecules carries, in addition to its oligonucleotide identifier, O, an oligonucleotide sequence, C, that is complementary to a corresponding oligonucleotide sequence carried by another library, B, of encoded bi-functional molecules. The two libraries are hybridised, thus creating functionally divalent bi-functional molecules that can be used in screening for targets with two binding sites. If appropriately designed, trimers may be formed instead of dimers, thus creating a library of functionally trivalent encoded molecules.

Figure 15. Molecule fragments used in example 1.

A). Molecule fragments employed in example X1 are shown.
B). List of the molecule fragments used at positions 0, 1, 2, and 3 in the library generation process of example X1.

5

EXAMPLES

Example 1:

Formation of a library of bifunctional molecules and affinity selection against the protein target Integrin $\alpha_v\beta_3$ receptor, employing subprocesses 3, 5, A and 1 (first synthesis-selection-amplification round), and A and 1 (second synthesis-selection-round) (see above), using amine acylations for the coupling of molecule fragments to generate the encoded molecules.

The human Integrin receptor $\alpha_v\beta_3$ is implicated in many biological functions such as inflammatory responses and thrombus formation as well as cellular migration and metastatic dissemination. The natural ligands for $\alpha_v\beta_3$ Integrin receptor contain an RGD tri-peptide consensus motif that interacts with the receptor binding pocket. Consequently, much medical research have focused on the synthesis and identification of small molecule RGD-mimetics with increased affinity for the $\alpha_v\beta_3$ receptor. One mimetic, Feuston 5 (Feuston *et al.*, *J Med Chem.* 2002 Dec 19;45(26):5640-8.), comprising an arginine bisostere coupled to a GD dipeptide exhibits a ten-fold increased affinity ($K_D = 111$ nM) compared to the RGD-tripeptide.

It would therefore be of interest to synthesize libraries of bifunctional molecules that include the molecule fragments that generate the Feuston 5 ligand. In the following protocols for the generation and screening of such libraries are described. First, the formation and screening of a 625-membered library is described.

Stage 1 synthesis: Generation of two sets of carriers, using chemical ligation and enzymatic ligation, respectively, during stage 1 synthesis to generate carrier molecules (subprocesses 3 and 5).

Figure 15 shows the molecule fragments and oligonucleotides employed to generate the library.

Formation of carrier molecules, Set 1: Five 14 nt oligonucleotides, each containing a 5'-terminal amino-group (Glen Research catalog # 10-1905-90) linked by a Spacer-PEG18

(Glen Research catalog # 10-1918-90) are synthesised by standard phosphoramidite chemistry, to give the following oligonucleotides:

- 5 O-0.1: 5'-NH₂-PEG-ATGCTCGAGCGCG-3'
 O-0.2: 5'-NH₂-PEG-TAGCTGTAGCGCG-3'
 O-0.3: 5'-NH₂-PEG-AGAGCTCTGACGCG-3'
 O-0.4: 5'-NH₂-PEG-CGTCTGTCTGACGCG-3'
 O-0.5: 5'-NH₂-PEG-ATGCTCGAGCGCG-3'

The sequences of these oligonucleotides are not crucial, and the sequences can be changed to increase the sequence dissimilarity or decrease the differences in annealing temperature.

- 10 Each of the O-0.n oligonucleotides (position 0 in the library) are now portioned out into separate wells (i.e., each oligonucleotide is placed in a separate well, here an eppendorf tube), and loaded with a specific molecule fragment, each of which comprises a carboxylic acid and a penteneoyl-protected amine. The five molecule fragments are shown in figure 15; one of these molecule fragments is penteneoyl-Asp(OMe)-OH (aspartic acid, where the side chain carboxylic acid has been protected with a methyl ester). The following molecule fragment loading protocol, Protocol A, is used:

- 15 1 nmol amino-modified oligonucleotide is lyophilized and then dissolved in 20 microliter of 100 mM Na-borate buffer, pH 8.0 with 90 mM sulfo-N-Hydroxysuccinimide (SNHS, Merck). The molecule fragments are preactivated by incubation of 15 microliter of 100 mM molecule fragment in DMSO and 15 microliter of 100 mM 1-ethyl-3-[3-

- 20 dimethylaminopropyl]carbodiimide hydrochloride (EDC, Merck) in DMF for 30 min at 30°C before addition to the oligonucleotide solution. Each of the five molecule fragments are added to a specific oligonucleotide, as described in figure 15. Following incubation for 45 min at 30°C, an additional 30 microliter of pre-activated molecule fragment is added and the solution incubated for another 45 min at 30 °C. Excess molecule fragment, activation agents, solvents and salt is removed by double gel filtration using Bio-rad microspin

- 25 columns 6 and eluted in MS-grade H₂O. Loading is optionally verified by Electrospray-MS analysis. Subsequently, the amino-protection group is removed by addition of 0.2 volumes of 25 mM iodine in a mixture of THF/H₂O (1:1) and incubated at 37 °C for 2 h. Excess iodine is quenched by 20 mM 2-mercaptoethanol before gel filtration purification using Bio-rad 6 microspin columns. From MS-analysis the efficiency of loading and deprotection can optionally be estimated. At the end of this first round of synthesis, each of the 5 oligonucleotides should be attached to their specific molecule fragment; the molecule fragment contains a reactive group, the amine, ready for reaction with the next molecule fragment that is added. The contents of the five wells are pooled, and redistributed into

- 35 five new wells.

Next, 1.2 nmoles unit identifier oligonucleotides, corresponding to position 1, are added to each well, according to the scheme of figure 15. The five oligonucleotides carry an

Imidazole-activated 5'-phosphate (Vischer, J., Schwarz, A.W. Journal of Molecular Evolution (1988), 28: 3-6; Zhao, Y., Thorson, J.S. Journal of Organic Chemistry (1998), 63:7568-7572) and have the following sequences:

- 5 O-1.1: 5'-Imp-CACAGGTACGAGG-3'
 O-1.2: 5'-Imp-CACATGATCTCTC-3'
 O-1.3: 5'-Imp-CATACATCTCTC-3'
 O-1.4: 5'-Imp-CATCTCTCTCTCA-3'
 O-1.5: 5'-Imp-CACAGCTCTCTCA-3'

10

- 1-2 nmoles of the oligonucleotide 3'-GCGGTGT-5' is added to all wells, and an appropriate buffer of pH 8-10 is added, to a final volume of 20-50 microliter. This oligo is complementary to the ends of oligonucleotides O-0.n and O-1.n, and by hybridization of these complementary sequences the 3'-OH group of the oligonucleotides O-0.n and the imidazole activated 5'-phosphates of the O-1.n oligonucleotides will be juxtaposed. The solution is incubated for 1-5 hrs at 37 °C or 50 °C. This results in ligation of the juxtaposed oligonucleotides, by formation of a phosphodiester bond.

- Optionally, the five solutions containing the ligation products are purified individually using a Bio-rad 6 spin columns according to manufacturer's instructions and lyophilized. Next, a specific molecule fragment is reacted with each of the five solutions of nascent bifunctional molecules, according to the scheme shown in Fig. 18, using loading protocol A described above. Excess free reactant, reagents and buffer is removed by gel filtration. The eluates are pooled, lyophilized and resuspended in 40 ul of H₂O before addition of 10 ul of 25 mM iodine (in THF/H₂O, ratio 1:1) for deprotection. The reaction is incubated at 37°C for 2 h. Excess iodine is quenched by addition of 1 ul of 1 M 2-mercaptoethanol and left at ambient temperature for 5 min before purification of the sample using spin-gel filtration (Bio-rad 6). The solution now contains 25 carrier molecules, where 25 different carrier identifier oligonucleotides each is attached to a specific one of 25 different dimers of molecule fragments. The carriers contain a free amino group, for reaction in the templated synthesis

- 30 (see below).

Formation of carrier molecules, Set II: The five 15 nt oligonucleotides, corresponding to position 2 of the library:

- O-2.1: 3'-SH-GACAGGACCCAG-5' P
 O-2.2: 3'-SH-CTCAGACCTACCAG-5' P
 35 O-2.3: 3'-SH-CGTGCTCTTACCAG-5' P
 O-2.4: 3'-SH-CCTGTGTCCAGCCAG-5' P
 O-2.5: 3'-SH-CTCAGAGGACCCAG-5' P

each carrying a 3'-terminal thiol-group, linked to the oligonucleotide through a flexible linker, and a 5'-terminal phosphate group, and each portioned out into one of five separate wells, are each linked through a thioester bond to a specific one of the five molecule fragments listed in figure 15 by the following Protocol B (Bruick et al., (1996)). Current biology 3:49-56):

Five N-protected molecule fragments (see figure 15) carrying a free carboxylic acid are first converted by standard procedures to the corresponding thioacids. After lyophilization, 1.2 equivalents of Ellmanns Reagent (5,5'-dithiobis(2-nitrobenzoic acid)) is incubated with the thioacid at pH 6.5 for 1 h, to produce the corresponding 5-thio-2-nitrobenzoic acid ester.

Optionally, the desired compounds are purified and characterized by HPLC and mass spectrometry.

1 nmol of each of the five oligonucleotides O-2.n are now incubated in separate wells with an excess of one of the five 5-thio-2-nitrobenzoic acid esters, according to the scheme of figure 15, at 25 °C or 37 °C, at pH 8 for 1-5 h. Optionally, 2 mM spermidine may be added to improve the efficiency of the reaction. Optionally, the formation of the correct oligonucleotide-thioester-molecule fragment product can be verified by mass spectrometry. Finally, the five modified oligonucleotides are pooled.

Excess molecule fragment, activation agents, solvents and salt is removed by double gel filtration using Bio-rad microspin columns 6 and eluted in MS-grade H₂O. Subsequently, the amino-protection group is removed by addition of 0.2 volumes of 25 mM iodine in a mixture of THF/H₂O (1:1) and incubated at 37 °C for 2 h. Excess iodine is quenched by 20 mM 2-mercaptoethanol before gel filtration purification using Bio-rad 6 microspin columns.

Alternatively, the oligonucleotides are precipitated with ethanol to remove the iodine. From MS-analysis the efficiency of loading and deprotection can optionally be estimated. At the end of this first round of synthesis, each of the 5 oligonucleotides should be attached to their specific molecule fragment through a thioester bond; the molecule fragment contains a free amine, ready for reaction with the next molecule fragment that is added. The contents of the five wells are pooled, and redistributed into five new wells.

Next, 1.2 nmoles unit identifier oligonucleotides, corresponding to position 3, are added to the wells, according to the scheme of figure 15. The five oligonucleotides have the following sequences:

O-3.1: 3' -CCTTAGTACGACG-5'
O-3.2: 3' -CCTTACGACGACG-5'
O-3.3: 3' -CCTGCTACTACTGCT-5'
O-3.4: 3' -CCTGCAATTCGAC-5'
O-3.5: 3' -CCTTATACGACG-5'

1-2 nmoles of the oligonucleotide 5'-TGGTCGAA-3', complementary to the ends of oligonucleotides O-2.n and O-3.n, is added to all wells. Then, the oligos are ligated in a volume of 20 ul using ligation buffer [30 mM Tris-HCl (pH 7.9), 10 mM MgCl₂, 10 mM DTT, 1 mM ATP] and 10 units T4-DNA ligase at ambient temperature for 1 hour. Subsequently, the 5 solutions of ligation products are purified individually using Biorad 6 spin columns, and the oligonucleotides lyophilized.

Next, a specific molecule fragment is reacted with the nascent bifunctional molecule according to the scheme shown in Fig.18 using loading protocol A described above. Excess free molecule fragment, reagents and buffer are then removed by gel filtration. The eluate is pooled, lyophilized and resuspended in 40 ul H₂O.

The BB-F3 molecule fragment does not react efficiently using protocol A, due to poor solubility of BB-F3 in organic solvent. Consequently, BB-F3 is reacted using Protocol C instead: The ligated and lyophilized sample is dissolved in 35 microliter 100 mM Na-borate buffer (pH 8.0) before addition of 10 microliter 100 mM BB-F3 in water and 5 microliter of 500 mM 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM, carboxylic acid activator) and incubated at 25 °C for 2 h. Following the coupling reaction, excess molecule fragment, reagent and salt is removed by gel filtration as described in protocol A.

The two solutions are pooled, and the resulting solution now contains 25 carrier molecules, all of which contain a thioester bond linking the molecule fragments to the carrier identifiers.

Stage 2 synthesis: Generation of bifunctional molecules by DNA-templated synthesis as described by (Bruick et al., (1996), Current biology 3:49-56) (subprocess A, see above).

Single-stranded DNA template library generation. First, four sets of duplex DNAs with overhangs are produced, by standard oligonucleotide synthesis followed by hybridization of appropriate oligonucleotide pairs, corresponding to the four encoded positions in the library. Each set of duplex DNA in this library contains five different dsDNAs, corresponding to the 5 different identifier sequences at each position, encoding 5 different molecule fragments at each positions. All dsDNA O.n carry a biotin as indicated below for the dsDNA O.1. When employing more molecule fragments, the number of dsDNAs in each set must be increased accordingly. An example of dsDNAs is shown below for the O-0.1, O-1.1, O-2.1, and O-3.1 identifiers:

dsDNA O.1.1:
5' -ATGCTCGAGACGCG-3'
3' -GTACGAGCTCT-5'

dsDNA 1.1:

5' -CACAGTAGACAGCATATGGCTTGGCCAAACACTG-3'
3' -GCGCGTGTTCATGCTTGACATACGACACCGGTTGTGAC-5'

5 dsDNA 2.1:

5' -GACCACAGGACGAGC-3'
3' -AAGGCTGGTGGTCTGCTC-5'

dsDNA 3.1:

5' -TTGGTTGACTAGACGAGCGCAGCATGATTC-3'
3' -AACCAACTGATCTCTGCTCGCTGCTACT-5'

Underlined sequences are priming sites for PCR amplification. All 5'-ends are

phosphorylated (contain phosphates). Overhang sequences can be extended in order to

allow more efficient ligation in the preparation of the templates described below. The X in

dsDNA 1.1 denotes a T that carries a biotin group.

The four sets of dsDNAs are incubated, to allow for hybridization between overhang, and

ligated as a mixture using ligation buffer [30 mM Tris-HCl (pH 7.9), 10 mM MgCl₂, 10 mM

DTT, 1 mM ATP] and T4-DNA ligase at ambient temperature for 1 hour. Thus, a total of

20 5x5x5=625 templates are generated. As an example, we have aligned the four dsDNAs

corresponding to the O-0.1, O-1.1, O-2.1, and O-3.1 identifiers immediately below; open

spaces highlight the complementary overhangs that hybridize during the ligation reaction:

25 dsDNA 3.1 dsDNA 2.1 dsDNA 0.1 dsDNA 1.1
-TTGGTTGACTAGACGAGCGCAGCATGATTC GACCACAGGACGAGC ATGCTCGAGACGCG CACAGTAGACAGCATATGGCTTGGCCAAACACTG-3
-AACCAACTGATCTCTGCTCGCTGCTACT AAGGCTGGTGGTCTGCTC GTACAGCTCT GCGCGTGTTCATGCTTGACATACGACACCGGTTGTGAC-5

30 The ligation product (the template) that results from the ligation of the above sequences is indicated immediately below:

dsDNA 3.1 dsDNA 2.1 dsDNA 0.1 dsDNA 1.1
5' -TTGGTTGACTAGACGAGCGCAGCATGATTCGACACGACGATCTCTGAGAGCGGCACAGTAGACAGCATATGGCTTGGCCAAACACTG-3'
3' -AACCAACTGATCTCTGCTCGCTGCTACTAAGGCTGGTGGTCTCTGCTGACAGCTCTGCGCTGTCTGCTATACGACACCGGTTGTGAC-5'

Thus, the sequence of identifier sequences in the 625 templates is: (primer annealing

site) - O-3 - O-2 - O-0 - O-1 - (primer annealing site). Optionally, more copies of the

40 templates can be produced by PCR amplification using primers that anneal to underlined

sequences. One of the primers should carry a T with biotin, as indicated. Thus, the ligation product and the PCR product contains a biotinylated lower strand.

The biotinylated double stranded product is now incubated with streptavidin-coated beads,

and the upper strand removed by alkaline denaturation of the strands, and the pH is

5 neutralized with an appropriate buffer to produce immobilized, single stranded template.

Hybridisation of carriers and template, and templated reaction between reactive groups

(Brulck et al., (1996), Current biology 3:49-56): 1-100 pmol immobilized templates are

mixed with an excess of Set I and Set II carriers obtained above, in an appropriate buffer

10 at pH 8. Optionally, the temperature is kept at 50 °C for 2 min., then lowered to 37 °C.

The temperature is now kept at 37 °C for 24 h. The carrier molecules anneal to the

template under these conditions; the close proximity of the amino group of carriers of Set

1, and the thioester of carriers of Set II, leads to amide formation, in effect transferring the

molecule fragment of the thioester carrier onto the amine carrier. At this point, 625

15 different bifunctional molecules have been generated.

Optionally, a DNA loop that can be ligated to the carrier molecule with the 0.n and 1.n

identifiers (to the right in the figure above) and to the template, and thus covalently

attaches the carrier molecule to the template, can be added. Thus, optionally, to the

20 bifunctional molecule that results from the templated synthesis immediately above, add an

oligonucleotide with the sequence 5'-TATGCGTTGGCCAAACACTGCGCAGATA-GAGGICTGCG-3',

where the stem sequences are underlined, and where the 5'-terminus is phosphorylated

(carries a phosphate). Add ligation buffer [30 mM Tris-HCl (pH 7.9), 10 mM MgCl₂, 10 mM

DTT, 1 mM ATP] and T4-DNA ligase at ambient temperature for 1 hour, to covalently attach

25 the rightward carrier molecule (carrying the encoded molecule that results from the

templated synthesis) to the template.

Optional amine and carboxylic acid deprotection. Optionally, to the solution of the previous

step is now added 0.2 volumes of 25 mM iodine (in THF/H₂O, ratio 1:1) for deprotection of

the penteneoyl-protected amines. Excess iodine is quenched by addition of 1 ul of 1 M 2-

30 mercaptoethanol and left at ambient temperature for 5 min before optional purification of

the sample using spin-gel filtration (Bio-rad 6). Then, optionally, NaOH is added to 25 mM,

at 80 °C for 5 minutes, to deprotect methyl/ester-protected carboxylic acids. Then increase

pH to 12.5 for one min. Optionally, the sample is purified using spin-gel filtration (Bio-rad 6).

35 Selection on immobilised target (subprocess 1).

Immobilisation and selection: Maxisorp ELISA wells (NUNC A/S, Denmark) is coated with

each 100 microliter 2 ug/mL Integrin alphaV/beta3 (Bachem) in PBS buffer [2.8 mM NaH₂PO₄,

7.2 mM Na₂HPO₄, 0.15 M NaCl, pH 7.2] overnight at 4°C. Then the integrin solution is

substituted for 200 microliter blocking buffer [TBS, 0.05% Tween 20 (Sigma P-9416), 1%

bovine serum albumin (Sigma A-7030), 1 mM MnCl₂ and incubated for 1 hour at room temperature. Then the wells are washed 2 times with 250 microliter blocking buffer, and 200 microliter blocking buffer, containing the library of bifunctional molecules generated above, is added to the wells. Following 2 hours incubation at room temperature the wells are washed with 20 x 250 microliter blocking buffer. After the final wash the wells are cleared with washing buffer and the bound bifunctional molecules eluted with MeOH, glycine pH 5, or an appropriate buffer of pH 11-13. The pH is adjusted to 7. The eluted fraction contains potential integrin alphaV/beta3 receptor ligands.

10 PCR amplification of the DNA templates of the isolated bifunctional molecules, and cloning and characterization: The templates of the eluted fraction is now amplified by PCR, and then either cloned and sequenced for characterization, or is taken through one more round of single-stranded template preparation, and stage 2 synthesis. For characterization, 5 ul eluted bifunctional molecules are used for PCR in a 25 ul reaction using 10ul Eppendorf hotmastermix 2.5x and 10 pmol each of forward and backwards primers that anneal to the underlined sequences depicted above. The PCR product is then ligated into suitable plasmid and transformed into e.g. E. coli, whereafter individual clones are sequenced by standard means (see for example below). From the DNA sequences the identity of the recovered encoded molecules can be deduced.

20 Template amplification, single-stranded template preparation, stage 2 synthesis (e.g. subprocess 5) and selection (e.g. subprocess 1). Instead of amplifying the recovered identifiers from the selection step above, and cloning and sequencing, the bifunctional molecules can be amplified and taken through one more round of selection. To this end, amplify the recovered identifiers with forwards and backwards primers, where the backwards primer carries a biotin (as indicated above). Isolate single-stranded DNA-template, add carriers generated above, and perform stage 2 synthesis as indicated above. Finally, the selection is performed, as indicated above, or by any other means that lead to identification of integrin alphaV/beta3 ligands. Finally, the identifiers recovered are PCR amplified, cloned, and sequenced (see for example below), to reveal the identity of the encoded molecules responsible for binding to the integrin receptor.

Identification and characterisation.

To obtain the sequences of the DNA templates, and thereby deduce the chemical structure of the encoded molecules, the double stranded PCR-product is cloned into e.g. an E.coli vector, propagated in E. coli, and individual clones sequenced. Each of the clones represent an identifier sequence of a bifunctional molecule in the pool isolated by the selections; from the the sequence of the DNA the corresponding encoded molecule (that was attached to the identifier of the same bifunctional molecule) can be deduced. The

TOPO-TA (Invitrogen Cat#K4575-J10) ligation is reacted with 4 ul PCR product, 1 ul salt solution, 1 ul vector. The reaction is incubated at RT for 30 min. Heat-shock competent TOP10 E.coli cells are thawed and put on ice. 5 ul ligation reaction is added. Following 30 min on ice, the cells are heat-shocked at 42°C water for 30 sec, then put on ice. 250 ul SOC is added and the cells incubated 1 h at 37°C, before spreading on LB-ampicillin plates followed by incubation ON at 37°C. Individual *E.coli* clones are picked and transferred to PCR wells containing 50 ul water. Colonies are incubated at 94°C for 5 minutes and 20 ul is used in a 25 ul PCR reaction with 5 pmol of each TOPO primer M13 forward & M13 reverse and Ready-To-Go PCR beads (Amersham) using the following PCR program: 94°C 2 min, then 30 x (94°C 4 sec, 50°C 30 sec, 72°C 1 min) then 72°C 10 min. Primers and free nucleotides are degraded by adding 1 ul EXO/SAP mixture 1:1 to 2 ul PCR product. Incubation is at 37°C for 15 min and then 80°C for 15 min. 5 pmol T7 primer is added and water to 12 ul. Subsequently, 8 ul DYE-namic ET cycle sequencing Terminator Mix is added followed by PCR-cycling using 30 rounds of (95°C 20 sec, 50°C 15 sec, 60°C 1 min).

15 Purification is done using seq96 spinplates (Amersham), followed by analysis on a MegaBace sequencer.

To verify that the isolated encoded molecules indeed represent ligands to the target protein (Integrin alphaV/beta3), individual bifunctional molecules may be prepared, by preparation of single stranded DNA of that bifunctional molecule, and performing the templated synthesis, to generate multiple copies of that specific bifunctional molecule. The ability of the bifunctional molecule (and, expectably, the ability of the encoded molecule) to bind the protein target (integrin alphaV/beta3) is then tested by e.g. immobilising the protein target in the well of a microtiter plate, adding the bifunctional molecule, washing off unbound bifunctional molecule, and then determine the amount of bound bifunctional molecule.

Alternatively, the identified encoded molecule may be synthesized in its free form, by standard chemical synthesis protocols, and then examined in e.g. competition binding experiments.

The directionality of the oligonucleotides used in the example may be changed, so as for example to include a thiol at the 5'-end rather than the 3'-end, or the sequences of the oligonucleotides may be changed in order to obtain highest possible mismatch ("sequence difference") among the different unit identifiers and carrier identifiers, while keeping the annealing temperatures relatively similar. This will increase the fidelity of the hybridization of carriers to the template during stage 2 synthesis, and will also increase the fidelity of the deconvolution step, since sequencing errors will be less of a problem if the identifiers have fewer identical nucleotide positions.

In the example a thioester was employed as the reactive group of Set II carriers. The activated ester can be any other type of activated ester (e.g., N-hydroxide succinimide ester, nitrophenyl-ester, nitrobenzyl-ester), or the ester may be a regular carboxyester. These activated esters are prepared by standard organic synthesis methods.

5

In the example, only the Set I carriers contain a long, flexible PEG linker. It may be advantageous that both carrier sets contain a PEG linker, to obtain high flexibility of the molecule fragments that must react.

10 In the example, the order of reactions between molecule fragments, and ligation of identifiers during stage 1 synthesis, is "reaction-ligation-reaction". This order can be changed, to be reaction-reaction-ligation, if desired.

The constant regions of the unit identifier oligonucleotides are 4 or 5 nt in the example.

15 The constant regions are complementary to the third oligonucleotide added; the third oligonucleotide brings the two unit identifiers into close proximity, and thus mediates the ligation of the unit identifiers. The overlap region between the identifier and the third oligonucleotide can be extended (to allow for a more efficient ligation during stage 1 synthesis), or shortened (to allow for more specific annealing of the carrier molecule during the stage 2 synthesis that follows; annealing is more specific because the sequence similarity with other carriers employed during the stage 2 templated synthesis will be smaller when the constant regions are shorter).

25 The recovered sequences from the selection experiment of example X1 will contain an abundance of the identifier sequences encoding the molecule fragments BB98, BB99, and BB-F3, as these are the molecule fragments that generate the known integrin alphaV/beta3 receptor ligand, Feuston 5.

30 The stage 1 synthesis protocol, stage 2 synthesis protocol, screening protocol, and characterization protocol, can be employed as modular units, as long as each of the four protocols are finalized by a purification to remove salts, reagents, unreacted molecule fragments, and the like. Often, an appropriate purification is spin-gel filtration (Bio-rad 6); in order to obtain very efficient purification, two spin-gel filtrations may be performed.

35 The following examples describe protocols for individual stage 1 synthesis, stage 2 synthesis, screening/selection, and characterization. As mentioned, these may be combined in any desired way, as long as each of the protocols are finalized with an appropriate purification step. Obviously, the length and composition of the identifiers must

be designed so as to mediate specific and efficient annealing of the carriers to the template during templated synthesis.

Example 2:

5 **Formation of five different libraries of bifunctional molecules, i.e., libraries containing 1.6×10^5 , 6.25×10^6 , 10^8 , or 10^{12} bi-functional molecules and affinity selection against the protein target Integrin alphaV/beta3 receptor, employing subprocesses 3), 5), A) and I), using amine acylations for the coupling of molecule fragments to generate the encoded molecules.**

10

This example describes the generation of libraries of five different libraries, i.e., libraries of 1.6×10^5 , 6.25×10^6 , 10^8 , or 10^{12} bi-functional molecules, and the use of these libraries for selection against the integrin alphaV/beta3 receptor.

The protocol described in example X1 is followed, except that the sets of molecule fragments are now changed so as to include 2, 20, 50, 100, or 1000 molecule fragments at each of the four positions, leading to the formation of libraries of $2 \times 2 \times 2 \times 2 = 16$, $20 \times 20 \times 20 \times 20 = 1.6 \times 10^5$, $50 \times 50 \times 50 \times 50 = 6.25 \times 10^6$, $100 \times 100 \times 100 \times 100 = 10^8$, or $1000 \times 1000 \times 1000 \times 1000 = 10^{12}$ bifunctional molecules. The molecule fragments carry the same N-protecting group (N-penteneoyl) and a free carboxylic acid, wherefore the protocol described in example X1 can be used, except that an appropriate number of wells are used, corresponding to the number of molecule fragments. A number of unit identifier oligonucleotides are used that correspond to the number of molecule fragments. Because of the size of these libraries, novel ligands not strongly related to the Feuston 5 ligand, will be identified from the bigger libraries. This is particularly true for the libraries of 10^8 or 10^{12} bifunctional molecules. For library sizes larger than 10^8 encoded molecules, ligands will be identified that do not contain all three molecule fragments BB98, BB99, and BB-F3, yet have dissociation constants lower than 100 micromolar.

Example 3:

30

Covalent attachment of a carrier to the template employed in the stage 2 synthesis.

The structure of the identifier template of the bifunctional molecule generated by stage 2 synthesis, and employed in the selections, can be varied. For example, before, during or 35 after the templated reaction, one of the carriers may be ligated to the template by a DNA ligase, if the template for example loops back on itself, as described in figure 4, example

3. Optionally, an extension reaction involving a primer that anneals to the other end of the template may be performed, in order to generate a duplex DNA where the encoded molecule is displayed at the end of the dsDNA. This may be done by annealing 1 nmol of a primer that is complementary to the end of the template that is not looping back on itself, and adding sequenase buffer containing 200 micromolar deoxy-ribonucleotides (dNTP) in a total volume of 100 microliter before addition of 20 units of sequenase and incubation at 30 °C for 1 h. Following extension the reaction mixture is used in the selection step without further purification.

10 Example 4:

Disulfide formation during stage 1 synthesis, employed to attach a scaffold molecule fragment comprising three reactive groups.

This is an example of a reaction that attaches a molecule fragment to another molecule fragment, or to the linker molecule L, through formation of a disulfide bond (Freskgård et al., WO 2004/039825 A2, example 1, p. 106-108). The protocol may be used in stage 1 synthesis. Similar reaction conditions can be employed in a stage 2 synthesis.

An amino-modifier C6 5'-labeled oligo (5'-X-CGTAACGACTGAATGACGT-3'), wherein X may be obtained from Glen research, cat. # 10-1039-90) was loaded with a peptide (Cys-Phe-Phe-Lys-Lys-Lys, CFFKKK) using SPDP activation (see below). The SPDP-activation of amino-oligo was performed using 160 µl of 10 nmol oligo in 100 mM Hepes-KOH, pH=7.5, and 40 µl 20 mM SPDP and incubation for 2 h at 30°C. The activated amino-oligo was extracted 3 times with 500 µl EtOAc, dried for 10 min in a speedvac and purified using micro bio-spin column equilibrated with 100 mM Hepes-KOH. The loading of peptide was then performed by adding 10 µl of 100 mM attachment entity and incubating overnight at 30°C. The loaded identifier oligo was precipitated with 2 M NH₄OAc and 2 volume 96% ethanol for 15 min at 80°C and then centrifuged for 15 min at 4°C and 15.000g. The pellet was re-suspended in water and the precipitation was repeated. Wash of the oligo-pellet was done by adding 100 µl of 70% ethanol and then briefly centrifuged. The oligo was re-dissolved in 50 µl H₂O and analysed by MS. After incubation the resin was removed by centrifugation and 15 µl of the supernatant was mixed with 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 µl acetonitrile. The sample was analysed using a mass spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The observed mass was 7244.93 Da, which correspond well with the calculated mass, 7244.00 Da. This experimental data exemplify the possibility to load a molecule fragment onto oligonucleotides through the formation of a disulfide bond. This particular molecule fragment (peptide) harbours three reactive groups, i.e. the amine groups of the lysine side chains, and therefore represents a scaffold with the ability to be reacted with one, two, or

three other molecule fragments that are capable of reacting with the amine groups (e.g. carboxylic acids).

Example 5.

Stage 1 acylation reaction.

This is an example of a stage 1 acylation reaction that attaches a molecule fragment to another molecule fragment coupled to an oligonucleotide, or to a reactive group of an oligonucleotide. Similar conditions can be applied for a stage 2 acylation reaction, except that the incoming molecule fragment must be at high concentration, e.g. 10-100 mM. The experiment is described in (Freskgård et al., WO 2004/039825 A2, p.129-137).

EDC-based acylation protocol: 10 µl triethanolamine (TEA) (0.1 M in DMF) was mixed with 10 µl molecule fragment (here called "Building Block (BB)"). The building blocks that were tested all carry a carboxylic acid and a Pent-4-enal amine protecting group; the concentration of the building block was 0.1 M in DMSO. From this mixture 6.7 µl was taken and mixed with 3.3 µl EDC [1-Ethyl-3-(3-Dimethylaminopropyl) carbodiimide Hydrochloride] (0.1 M in DMF) and incubated 30 minutes at 25°C. 10 µl of the Building block-EDC-TEA mixture was added to 10 µl of an amino modified oligonucleotide (here termed "amino oligo") (in 0.1 M HEPES buffer ((4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid, SIGMA), pH 7.5 and incubated for 30 minutes. During this half hour, another 6.7 µl of BB-TEA mix was mixed with 3.3 µl EDC (0.1M in DMF) and incubated for 30 minutes at 25°C. 10 µl of this second BB-EDC-TEA mixture was then added to the amino oligo mixture together with 10 µl of 0.1 M HEPES buffer to maintain a 1:1 ratio of DMSO/DMF: H₂O. Then the mixture was incubated for 30 minutes. During this half hour, another 6.7 µl of BB-TEA mix was mixed with 3.3 µl EDC (0.1 M in DMF) and incubate for 30 minutes at 25°C. 10 µl of this third BB-EDC-TEA mixture was then added to the amino oligo mixture together with 10 µl of 0.1 M HEPES buffer to maintain a 1:1 ratio of DMSO/DMF: H₂O. Then the mixture was incubated for 30 minutes. The oligonucleotide, linked to the molecule fragment (here termed "loaded oligo") was then purified by gel filtration with columns (Biospin P-6, BioRad) equilibrated with water. The pent-4-enal amine protection group was then removed by addition of 0.25 volumes 25 mM I₂ in 1:1 water:tetrahydrofuran (THF) and incubation at 37°C for 2 hours. The mixture was then purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. Loaded oligos were analyzed by ES-MS. Molecule fragments tested included aliphatic as well as aromatic compounds, and all were attached efficiently through amide bond formation, as evidenced by mass spectrometric data within a few Daltons of the expected mass. See (Freskgård et al., WO 2004/039825 A2, p.129-137).

DMT-MM-based acylation protocol: 10-15 nmol of carrier oligo 2 was lyophilized and redissolved in 27.5 μ l H₂O. To this was added 7.5 μ l 1 M HEPES pH 7.5, 10 μ l of 2-amino-pent-4-enal protected (allyl-glycine) building block (0.1 M in dimethyl sulfoxide), and 5 μ l DMT-MM [4-(4,6-dimethoxy-1,3,5-thiazin-2-yl)-4-methylmorpholinium chloride] (0.5 M in water). The mixture was incubated 4-16 hours at 25-30°C. The oligo was purified by gel filtration (Biospin P-6, BioRad). To convert the methyl ester moiety of the building block to a carboxylic acid, 5 μ l 0.4 M NaOH was added and the mixture was incubated 20 min at 80°C. The mixture was then neutralized by adding 10 μ l 0.5 M HEPES pH 7.5 and 5 μ l 0.4 M HCl. The loaded building block oligo was purified by gel filtration Biospin P-6, BioRad) and analyzed by ES-MS. Aliphatic as well as aromatic building blocks were attached to the amine modified oligonucleotide efficiently, as evidenced by the MS-data which showed good correlation between expected and observed mass. See (Freskgård et al., WO 2004/039825 A2, p.129-137).

15

Example 6.

Stage 1 enzymatic ligation of oligonucleotides carrying molecule fragments.

This is an example of a stage 1 enzymatic ligation that attaches one oligonucleotide, carrying a molecule fragment, to another oligonucleotide through covalent phosphodiester bond formation. The experiment is described in (Freskgård et al., WO 2004/039825 A2, p.137-143).

500 pmol loaded carrier oligo (oligonucleotide carrying a molecule fragment), 5'-phosphorylated, was mixed with 750 pmol anti-codon oligo (not carrying any molecule fragment) and 750 pmol splint oligo (comprising complementary sequences to both the carrier oligo and the anti-codon oligo). See figure immediately below, showing an example pair of carrier oligo and anti-codon oligo, as well as the splint oligo. Note that the anti-codon oligo comprises Inosines (allowing annealing to several different bases, here C and A. The mixture was lyophilized and redissolved in 15 μ l water. Oligos were annealed by heating and slowly cooling to 20°C. 15 μ l TaKaRa ligase mixture (Takara Bio Inc) was added and the reaction was incubated at 20°C for 1 hour. The mixture was purified by gel filtration (Biospin P-6, BioRad) and the efficiency of the ligation was checked by running an aliquot on a Novex TBE-UREA gel (Invitrogen). Both oligonucleotides carrying aliphatic and aromatic compounds were tested; different sequences around the ligation point was examined as well. All oligonucleotides tested were ligated with more than 95% efficiency. See (Freskgård et al., WO 2004/039825 A2, p.137-143).

35

Loaded carrier Oligo	Anti-codon oligo
3'-2GGAGTCGACACATAGCTCGCG	CGTCGIIIIIIIGCAGCCATATAGTCGT-X
TCCAGCG--GCAGCCA	
Splint oligo	

5

Example X7. Stage 1 synthesis of a 484-member library of bifunctional molecules (subprocess 4), and selection by affinity selection on immobilized target (subprocess I).

This is an example of a stage 1 synthesis employing subprocess 4. Three rounds of encoding 10 are employed, involving 4, 11 and 11 molecule fragments, thus generating a total of 484 bifunctional molecules that may be used as carriers in a templated stage 2 synthesis. The three rounds of encoding involve acylation reactions. (Freskgård et al., WO 2004/039825 A2, p.143-148).

First encoding round. 2 pmol of loaded identifier oligo 1.1 (i.e, a particular molecule fragment attached to the identifier oligonucleotide) was combined with 200 pmol of each loaded identifier oligo 1.2, 1.3, and 1.4. (602 pmol loaded identifier oligos in total). These were mixed with 0.7 pmol building block oligo 3.1.3. (i.e., a particular molecule fragment attached to an oligonucleotide, capable of hybridizing with the identifier oligonucleotide), and 72.7 pmol each of 10 different other first round building block oligos (eg. 3.1.1 and 3.1.2; 727 pmol loaded building block oligos in total). The oligos were lyophilized and redissolved in 50 μ l extension buffer (EX) [20 mM HEPES, 150 mM NaCl, 8 mM MgCl₂]. The mixture was heated to 80°C and slowly cooled to 20°C to allow efficient annealing of identifier and building block oligos. 5 μ l of 0.5 M DMT-MM in water was added and the mixture was incubated at 37°C for 4 hours. Extension of the identifier oligo on the building block oligo identifier was performed by adding 3 μ l of a 10 mM mixture of each deoxynucleotide triphosphate [dATP, dGTP, dCTP, dTTP] and 3 μ l of 13 units/ μ l Sequenase (Amersham Biosciences). The mixture was subsequently incubated at 30°C overnight. Then 3 μ l of 2M NaOH was added and the mixture was incubated for 80°C for 10 minutes followed by neutralization by addition of 3 μ l 2M HCl. The mixture was then purified by passing through a gel filtration column (Biospin P-6, BioRad). 0.25 volumes of 25 mM I₂ in 1:1 THF:water was added, mixed and incubated at 37°C for 2 hours. 60 μ l binding buffer (BF) [100 mM HEPES, 150 mM NaCl] and water ad 300 μ l was added. The mixture was added to streptavidin-sepharose beads (Amersham Biosciences) pre-washed 3 times in BF buffer and incubated at room temperature for 10 minutes, followed by incubation on ice for 10 minutes with gentle stirring. The beads were then washed three times with water. Extended Identifier oligos were stripped from the building block oligos bound to the streptavidin-

35

sepharose beads by applying 100 ul NH3 1:1 in water and incubating at room temperature for 5 minutes.

Second encoding round. To the eluate was added 0.36 pmol second round loaded building block oligo 3.2.2 and 36.4 pmol each of 10 different other second round building block oligos (eg. 3.2.1 and 3.2.3; 364 pmol loaded second round building block oligos in total) and the mixture was lyophilized and redissolved in 50 ul EX buffer. The encoding was performed essentially as described under above.

Final extension. The eluted identifier oligo were lyophilized and dissolved in 50 ul EX buffer. Then 200 pmol primer E38 [5'-XTTTTAGTCGACAT-3\ X=CXS Biotin] was added. Annealing was performed by heating the mixture to 80°C and slowly cooling to 20°C. Extension of the identifier oligo was performed by adding 3 ul of a 10 mM mixture of each deoxynucleotide triphosphate [dATP, dGTP, dCTP, dTTP] and 3 ul of 13 units/ul Sequenase. The mixture was subsequently incubated at 30°C for 2 hours. The mixture was then purified by passing through a gel filtration column (Biospin P-6, BioRad). This eluate was used for selection. An aliquot was removed for analysis of the Input in the selection procedure.

General procedure 5: affinity selection on immobilized protein target. Maxisorp ELISA wells (NUNC A/S, Denmark) were coated with each 100 ul 2ug/mL Integrin α Vp3 (Bachem) in PBS buffer [2.8 mM NaH₂PO₄, 7.2 mM Na₂HPO₄, 0.15 M NaCl, pH 7.2] overnight at 4°C. Then the integrin solution was substituted for 200 pl blocking buffer [TBS, 0.05% Tween 20 (Sigma P-9416), 1% bovine serum albumin (Sigma A-7030), 1 mM MnCl₂] which was left on for 3 hours at room temperature. Then the wells were washed 10 times with blocking buffer and the encoded library was added to the wells after diluting it 100 times with blocking buffer. Following 2 hours incubation at room temperature the wells were washed 10 times with blocking buffer. After the final wash the wells were cleared of wash buffer and subsequently inverted and exposed to UV light at 300-350 nm for 30 seconds. Then 100 ul blocking buffer without Tween-20 was immediately added to each well, the wells were shaken for 30 seconds, and the solutions containing eluted identifiers were removed for PCR analysis.

Analysis of selection input and output. PCR was performed on the input and output of the selection, using primers corresponding to the 5' end of the identifier oligos and the E38 primer. PCR was performed using Ready-To-Go (RTG) PCR beads (Amersham Biosciences) and 10 pmol each primer in a reaction volume of 25 ul. The PCR reaction consisted of an initial denaturation step of 94°C for 2 minutes followed by 30-45 cycles of denaturation at 94°C for 30 seconds, annealing at 58°C for 1 minute and extension at 72°C for 1 minute. A final extension step of 2 minutes at 72°C was included. The PCR products were resolved by agarose gel electrophoresis and the band corresponding to the expected size was cut from the gel and purified using QIAquick Gel Extraction Kit (QIAGEN). To sequence individual PCR fragments the purified PCR products were cloned into the pCR4-TOPO vector (Invitrogen) according to the manufacturer's instructions. The resulting mixture was used

for transformation of TOP10 *E. coli* cells (Invitrogen) using standard procedures. The cells were plated on growth medium containing 100 ug/ml ampicillin and left at 37°C for 12-16 hours. Individual *E. coli* clones were picked and transferred to PCR wells containing 50 ul water. These wells were then boiled for 5 minutes and 20 ul mixture from each well was used in a PCR reaction using RTG PCR beads and 5 pmol each of M13 forward and reverse primers according to the manufacturer's instructions. A sample of each PCR product was then treated with Exonuclease I (USB) and Shrimp Alkaline Phosphatase (USB) to remove degrade single stranded DNA and dNTPs and sequenced using the DYEnamic ET cycle sequencing kit (Amersham Biosciences) according to the manufacturer's instructions and the reactions were analyzed on a MegaBace 4000 capillary sequencer (Amersham Biosciences). Sequence outputs were analyzed with ContigExpress software (Informa Inc.). A overview of molecule fragments used for library generation is shown in (Freskgård et al., WO 2004/039825 A2, p.146-147).

Theoretically, the integrin α Vp3 ligand A (Molecule 7 in Feuston B. P. et al., Journal of Medicinal Chemistry 2002, 45, 5640-5648) is present in 1 out of 3x10⁸ bifunctional molecules in this library. The codon combination compatible with encoding of ligand A was not found in 28 sequences derived from the encoded library before selection (input) in agreement with the expected low abundance of this codon combination (1 in 3x10⁸). A codon combination compatible with encoding of ligand A was found in 5 out of 19 sequences derived from the encoded library after selection in Integrin α VB3-coated wells. These numbers thus correspond to an apparent enrichment factor of $(3 \times 10^8 / (19 / 7)) = 8 \times 10^7$.

For more detailed data see (Freskgård et al., WO 2004/039825 A2, p.143-148).

Example 8. Selection of bifunctional molecules using size-exclusion chromatography.

This is an example of subprocess iii), although a real library of bifunctional molecules are not screened. A protocol for selection employing size-exclusion chromatography is presented. The experiment is taken from (Freskgård et al., WO 2004/039825 A2, p.148-150).

This example illustrates the possibility to use column separation to perform selection on complexes against various targets. In this example, size-exclusion chromatography (SEC) is used, but other types of chromatography can be used where target-bound complexes are separated from the non-bound complexes. The complex is exemplified in this example by a biotin molecule attached to an oligonucleotide sequence with a predetermined sequence (see below). Thus, the nucleotide sequence of the identifier specifies the identity of the

synthetic molecule as biotin. The encoding sequence can have any length and be divided into discrete regions for encoding various building blocks as discussed elsewhere herein. Also, the displayed molecule can have a linear or scaffold structure. Biotin-AATTCGGGAACACTAGTCAACATGA Biotin is known to bind to streptavidin. The binding of 5 biotin to streptavidin will link the identifier to the target molecule and therefore change the identifiers physical and chemical properties, such as e.g. the apparent molecular weight. This change is possible to detect using e.g. size-exclusion chromatography: 78 pmol of the complex molecule was loaded on a Superdex 200, PC 3.2/30 column (AKTA-FPLC, AmershamPharmaciaBiotech) and analysed in PBS buffer with a flow rate of 0.050 ml/min.

10 As may be seen from the spectrogram, the complex molecules retention-time was approximately 35 minutes. When the target (83 pmol streptavidin) was analysed under identical conditions the retention-time was approximately the same. The low absorption of the target molecules is due to the wavelength (260 nm) used in the measurement. At this wavelength, the extinction coefficient is high for the nucleotides in the complexes but low 15 for the protein target.

However, when the complex molecules was premixed with the target molecules (78 pmol complex and 83 pmol target incubated for about 1 h in PBS buffer) to allow binding and then analysed under identical conditions, the retention-time change significantly (28 20 minutes). The change is due to the increase in molecular weight (or hydrodynamic volume) due to the binding of the complex to the target. This will allow

the separation of the target-bound complexes from the non-bound complexes. The fraction that contains the complexes and the target molecules are pooled and amplified using appropriate primers. The amplified identifiers can then be used to decode the structures of 25 the enriched displayed molecules. The strategy of performing column-selection of libraries of bifunctional complexes has two major advantages. First, the enriched (target-bound) complexes are eluted before the non-bound complexes, which will drastically reduce the background from the non-bound complexes. Secondly, the enrichment on the column will be extensive due to all the separation steps in the pores in the matrix. The separation of 30 the target-bound complexes using this approach will be dependent on the molecular weight of the complexes but predominantly of the molecular weight of the target. The molecular weight of the target can be adjusted by linking the target to a support that increases the apparent molecular weight. The increased molecular weight will enhance the separation by reducing the retention-time on the column. This can be done using for example a fusion 35 protein, antibody, beads, or cross-linking the target in multimeric form. Thus, the target protein can be expressed as a fusion protein or a specific antibody can be used to increase the molecular weight. The target can be immobilized on small beads that permit separation and the target can be cross-linked using standard reagents to form multimers or cross-

linked to a carrier molecule, for example another protein. Preferably, the molecular weight is increased so the target molecules elute in the void volume of the column. Examples of other types of column separation that can be used are affinity chromatography, hydrophobic interaction chromatography (HIC), and ion-exchange chromatography. Examples of column media, other than Superdex, that can be used in 5 size-exclusion chromatography are: Sephacryl, Sepharose or Sephadex.

Example 9.

Encoded multiple component reaction (MCR) during a stage 1 synthesis.

- 10 This is an example of a stage 1 synthesis that involves the reaction of multiple different encoded molecule fragments in the same well; this is an example of an UGI reaction. The experiment is described in (Freskgård et al., WO 2004/039825 A2, p.157-162).
- Preparation of aldehyde-comprising scaffold-oligo, using 4-carboxybenzaldehyde. A solution of 4-carboxybenzaldehyde (scaffold) in DMF (25 μ L, 150 mM) was mixed with 25 μ L of a 150 mM solution of EDC in DMF. The mixture was left for 30 min at 25°C. 50 μ L aminooligo (10 nmol) in 100 mM HEPES buffer pH 7.5 was added and the reaction mixture was left for 20 min at 25°C. Excess scaffold was removed by extraction with EtOAc (500 μ L) and remaining EtOAc was removed *in vacuo* by spinning 10 min in a speedvac. The mixture was then 20 purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. The loaded oligo were analyzed by ES-MS.
- Multi-component reaction. A solution of the Benzaldehyde loaded oligo prepared above (200 pmol) was lyophilized and redissolved in 10 μ L H₂O. 2-Methoxy ethylamine in methanol (10 μ L, 40mM), 3-furan-2-yl-acrylic acid in methanol (10 μ L, 40 mM), and cyclohexyl isocyanide in methanol (10 μ L, 40mM) was added and incubated overnight at 37 °C. The reaction 25 mixture was diluted with 40 μ L H₂O and purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. MCR-product on oligo was analyzed by ES-MS. The starting benzaldehyde loaded oligo was identified in the MS-spectrum together with the UGI product.
- Multi-component reaction. A solution of benzaldehyde loaded oligo (320 pmol) was 30 lyophilized and redissolved in 10 μ L H₂O. 2-Amino ethanol in methanol (10 μ L, 40mM), 3-Methoxy-propionic acid in methanol (10 μ L, 40 mM), and ethyl isocynoacetate in methanol (10 μ L, 40mM) was added and incubated overnight at 37 °C. The reaction mixture was diluted with 40 μ L H₂O and purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. MCR-product on oligo was analyzed by ES-MS. The starting 35 benzaldehyde loaded oligo was identified in the MS-spectrum together with three products, Diketopiperazine, UGI product and the Amine product.

Encoding. Excess reactants, activation agents, solvents and salt was removed by double gel-filtration using Bio-rad microspin columns 6 and eluted in MS-grade H₂O and loading was verified by Electrospray-MS (Bruker Inc) analysis before the displayed molecule attached to the oligonucleotide was encoded. The benzaldehyde loaded oligonucleotide, that has been 5 reacted with the other three components to form the displayed molecule as described above was mixed with the codon oligonucleotides L2, L3 and L4 together with the splint oligonucleotides S1, S2 and S3 (sequences shown below) and ligated using a ligase (T4 DNA ligase). The ligation was performed using the following conditions. The double stranded oligonucleotide was achieved by mixing the encoding strands (L1, L2, L3 and L4) 10 with the splint oligonucleotides (S1, S2 and S3) to form a 7 oligonucleotide hybridisation product (for efficient annealing and ligation). About 50 pmol of each specific oligonucleotide was used and the oligonucleotides was ligated in a volume of 20 μ L using ligation buffer [30 mM Tris-HCl (pH 7.9), 10 mM MgCl₂, 10 mM DTT, 1 mM ATP] and 10 units T4-DNA ligase at ambient temperature for 1 hour.

15 L1: 5'-CGATGGTACGTCAGGTCGCA-3'
 S1: 5'-ATCGTGCTGCGACCT-3'
 L2: 5'-GCACGATATGTACGATACACTGA-3'
 S2: 5'-GTGCCATTCAAGTGT-3'
 L3: 5'-ATGGCACTTAATGGTTGTAATGC-3'
 S3: 5'-TGATATGCGCAATTAC-3'
 L4: 5'-GCATACAAATCGATAATGCAC-3'

The identifier comprising the tags was amplified using a forward (FP) and reverse (RP) primer using the following conditions: 5 μ L of the ligated identifier oligonucleotide was used for PCR in a 25 μ L reaction using 10 μ L Eppendorf hotmastermix 2.5x and 10 pmol each of 25 AH361 & Fw-27. PCR was run: (ENRICH30): 94°C 2 min, then 30 cycles of [94°C 30 sec, 58°C 1 min, 72°C 1 min], then 72°C 10 min.

FP: 5'-CGATGGTACGTCAGGTCGCA-3' RP: 5'-GTGCATTATCGATTGTATGC-3'

The amplified identifier oligonucleotide was cloned to verify that the assembled oligonucleotides contained the codon region (CGTCC, GTACG, AATGG and TCGAT). The 30 TOPO-TA (Invitrogen Cat#K4575-J10) ligation was reacted with 4 μ L PCR product, 1 μ L salt solution, 1 μ L vector. The reaction was incubated at RT for 30 min. Heat-shock competent TOP10 E.coli cells was thawed and put on ice. 5 μ L ligation reaction was added. Following 30 min on ice, the cells were heat-shocked at 42°C water for 30 sec, and then put on ice. 250 μ L SOC was added and the cells incubated 1 h at 37°C, before spreading on LB-ampicillin 35 plates followed by incubation ON at 37°C. Individual E.coli clones were picked and transferred to PCR wells containing 50 μ L water. Colonies were incubated at 94°C for 5 minutes and 20 μ L was used in a 25 μ L PCR reaction with 5 pmol of each TOPO primer M13

forward & M13 reverse (AH365/AH366) and Ready-To-Go PCR beads (Amersham) using PCR program: 94°C 2 min, then 30 x (94°C 4 sec, 50°C 30 sec, 72°C 1 min) then 72°C 10 min.

Primers and free nucleotides were degraded by adding 1 μ i EXO/SAP mixture 1:1 to 2 μ L 5 PCR product. Incubation was at 37°C for 15 min and then 80°C for 15 min. 5 pmol T7 primer (AH368) was added and water to 12 μ L. Subsequently, 8 μ L DYE-namic ET cycle sequencing Terminator Mix was added followed by PCR-cycling using 30 rounds of (95°C 20 sec, 50°C 15 sec, 60°C 1 min). Purification was done using seq96 spinplates (Amersham), followed by analysis on a MegaBace sequenizer.

Example 10.

Stage 1 "click" reaction.

This is an example of stage 1 synthesis, using the "click" reaction. Similar conditions can 15 be applied to stage 2 "click" reactions. The experiment is described in (US patent application 60/588,672, p. 34-35.)

General procedure. An alkyne-containing DNA conjugate is dissolved in pH 8.0 phosphate buffer at a concentration of ca. 1mM. To this mixture is added 10 equivalents of an organic azide and 5 equivalents each of copper (II) sulfate, ascorbic acid, and the ligand (tris-(1-benzyltriazol-4-yl)methyl)amine) all at room temperature. The reaction is followed by LCMS, and is usually complete after 1 ~ 2 h. The resulting triazole-DNA conjugate can be isolated by ethanol precipitation.

Preparation of Azidoacetyl-Gly-Pro-Phe-Pra-NH₂. Using 0.3 mmol of Rink-amide resin, the indicated sequence was synthesized by automated synthesis with Fmoc-protected amino 25 acids and HATU as activating agent (Pra = C-propargylglycine). Azidoacetic acid was used to cap the tetrapeptide. The peptide was cleaved from the resin with 20% TFA/DCM for 4 h. Purification by RP HPLC afforded product as a white solid (75 mg, 51%). ¹H NMR

(DMSO-d₆, 400 MHz): 8.4 - 7.8 (m, 3H), 7.4 - 7.1 (m, 7 H), 4.6 - 4.4 (m, 1H), 4.4 - 4.2 (m, 2H), 4.0 - 3.9 (m, 2H), 3.74 (dd, 1H, J = 6 Hz, 17 Hz), 3.5 - 3.3 (m, 2H), 3.07 (dt, 1H, J = 5 Hz, 14 Hz), 2.92 (dd, 1H, J = 5 Hz, 16 Hz), 2.86 (t, 1H, J = 2 Hz), 2.85 - 2.75 (m, 1H), 2.6-2.4 (m, 2H), 2.2 - 1.6 (m, 4H). IR (mul) 2900, 2100, 1450, 1300 cm⁻¹.

ESIMS 497.4 ([M+H], 100%), 993.4 ([2M+H], 50%). ESIMS with ion-source fragmentation: 519.3 ([M+Na], 100%), 491.3 (100%), 480.1 ([M-NH₂], 90%), 452.2 ([M-NH₂-CO], 20%), 424.2 (20%), 385.1 ([M-Pra], 50%), 357.1 ([M-Pra-CO], 40%), 238.0 35 ([M-Pra-Phe], 100%).

Cyclization of Azidoacetyl-Gly-Pro-Phe-Pra-NH₂: The azidoacetyl peptide (31 mg, 0.62 mmol) was dissolved in MeCN (30 mL). Diisopropylethylamine (DIEA, 1 mL) and Cu(MeCN)₄PF₆ (1 mg) were added. After stirring for 1.5 h, the solution was evaporated and the resulting residue was taken up in 20% MeCN/H₂O. After centrifugation to remove

Insoluble salts, the solution was subjected to preparative reverse phase HPLC. The desired cyclic peptide was isolated as a white solid (10 mg, 32%). ¹H NMR (DMSO-d₆, 400 MHz): 8.2S (t, 1H, J = 5 Hz), 7.77 (s, 1H), 7.2 - 6.9 (m, 9H), 4.98 (m, 2H), 4.48 (m, 1H), 4.28 (m, 1H), 4.1 - 3.9 (m, 2H), 3.63 (dd, 1H, J = 5 Hz, 16 Hz), 3.33 (m, 2H), 3.0 (m, 3H), 2.48 (dd, 1H, J = 11 Hz, 14 Hz), 1.75 (m, 1H), 1.55 (m, 1H), 1.32 (m, 1H), 1.05 (m, 1H). IR (mul) 2900, 1475, 1400 cm⁻¹. ESIMS 497.2 ([M+H], 100%), 993.2 ([2M+H], 30%), 1015.2 ([2M+Na], 15%). ESIMS with ion-source fragmentation: 535.2 (70%), 519.3 ([M+Na], 100%), 497.2 ([M+H], 80%), 480.1 ([M-NH₂], 30%), 452.2 ([M-NH₂-CO], 40%), 208.1 (60%).

10

Example 11:**A stage 1 synthesis Involving aromatic nucleophilic substitution.**

This is an example of an aromatic nucleophilic substitution reaction employed in a stage 1 synthesis. Similar conditions may be used in stage 2 synthesis. The experiments are described in (US patent application 60/588,672, p. 36)

General Procedure for Arylation of DNA-linker with Cyanuric Chloride: DNA-Linker is dissolved in pH 9.5 borate buffer at a concentration of 1 mM. The solution is cooled to 4° C and 20 equivalents of cyanuric chloride is then added as a 500 mM solution in MeCN. After 20 h, complete reaction is confirmed by LCMS and the resulting dichlorotriazine-DNA conjugate is isolated by ethanol precipitation.

Procedure for Amine Substitution of Dichlorotriazine-DNA: Dichlorotriazine-DNA is dissolved in pH 9.5 borate buffer at a concentration of 1 mM. At room temperature, 40 equivalents of an aliphatic amine is added as a DMF solution. The reaction is followed by LCMS and is usually complete after 2 h. The resulting monochlorotriazine-DNA conjugate is isolated by ethanol precipitation.

Procedure for Amine Substitution of Monochlorotriazine-DNA: (Alkylamino)-monochlorotriazine-DNA is dissolved in pH 9.5 borate buffer at a concentration of 1 mM. At 42° C, 40 equivalents of a second aliphatic amine is added as a DMF solution. The reaction is followed by LCMS and is usually complete after 2 h. The resulting diamino-triazine-DNA conjugate is isolated by ethanol precipitation.

Example 12:**35 A stage 1 synthesis (subprocess 9) and characterization of a library of 10⁵ members.**

This is an example of a stage 1 synthesis, involving five synthesis rounds (here termed "cycles"), employing acylation reactions for the coupling of molecule fragments (here

termed "building blocks"). Similar conditions can be used in stage 2 synthesis. The experiments are described in (US patent application 60/588,672, p. 26-34). The synthesis of a library comprising on the order of 10⁴ distinct members was accomplished using the following reagents:

5

Compound 1: An approximately 19 bp duplex DNA, where the two strands at one end has been covalently linked, and where that end includes a PEG linker and a terminal amino group; and where the 5'-end of one strand at the other end carries a 5'-phosphate.

Building block precursors: 12 compounds, each of which contains a Fmoc-protected amino 10 group and a free carboxylic acid. The compounds include aliphatic as well as aromatic compounds and aliphatic cyclic structures.

Oligonucleotide tags: A total of 60 duplex DNAs, with 7 central base pairs and 2 nt overhangs at both ends, and 5'-phosphates at both ends, are included. The 60 duplex DNAs correspond to 5 cycles where 12 tags are added, one per building block precursor, in 15 each round.

IX ligase buffer: 50 mM Tris, pH 7.5; 10 mM dithiothreitol; 10 mM MgCl₂; 2.5 mM ATP; 50 mM NaCl.

20 10X ligase buffer: 500 mM Tris, pH 7.5; 100 mM dithiothreitol; 100 mM MgCl₂; 25 mM ATP; 500 mM NaCl

Cycle 1

To each of twelve PCR tubes was added 50 µl of a 1 mM solution of Compound 1 in water; 25 75 µl of a 0.80 mM solution of one of Tags 1.1-1.12; 15 µl 10X ligase buffer and 10 µl deionized water. The tubes were heated to 95 °C for 1 minute and then cooled to 16 °C over 10 minutes. To each tube was added 5,000 units T4 DNA ligase (2.5 µl of a 2,000,000 unit/mL solution (New England Biolabs, Cat. No. M0202)) in 50 µl IX ligase buffer and the resulting solutions were incubated at 16 °C for 16 hours.

30 Following ligation, samples were transferred to 1.5 ml Eppendorf tubes and treated with 20 µl 5 M aqueous NaCl and 500 µl cold (-20 °C) ethanol, and held at -20 °C for 1 hour. Following centrifugation, the supernatant was removed and the pellet was washed with 70% aqueous ethanol at -20 °C. Each of the pellets was then dissolved in 150 µl of 150 mM sodium borate buffer, pH 9.4.

35 Stock solutions comprising one each of building block precursors BB1 to BB12, N,N-disopropylethanolamine and O-(7-azabenzotriazol-1-yl)-1,3,3-tetramethyluronium hexafluorophosphate, each at a concentration of 0.25 M, were prepared in sodium phosphate buffer, pH 8.0, and incubated at room temperature for 20 minutes. Each solution (6 µl) was diluted with 30 µl N₃N'-dimethylformamide and added to the

appropriate eppendorf tube. Two additional 6 μ L aliquots of building block precursor stock solution were added after 20 minutes and 40 minutes, respectively, for a final ratio of 30:1 building block precursor to tag. The tubes were gently shaken for 2 hours at 4°C. The tags and corresponding building block precursors used in Round 1 are set forth in Table 1, 5 below.

Table 1

Building Block Precursor	Tag
BB1	1.11
BB2	1.6
BB3	1.2
BB4	1-8
BB5	1.1
BB6	1.10
BB7	1.12
BB8	1.5
BB9	1.4
BB10	1.3
BB11	1.7
BB12	1.9

Following acylation, the 12 reaction mixtures were pooled and the resulting mixture was lyophilized to yield a dry residue, which was dissolved in 1.7 mL water. Two volumes of cold 100% ethanol were added and the mixture was allowed to stand at -20°C for at least 10 one hour. The mixture was then centrifuged for 15 minutes at 14,000 rpm in a 4°C microcentrifuge. Following centrifugation, as much supernatant as possible was removed with a 1 mL micropipet; the mixture was then centrifuged again, and the remainder of the supernatant was removed with a 200 μ L pipet. Cold 70% ethanol (200 μ L) was then added to the tube, and the mixture was centrifuged for 5 minutes at 4°C.

15 The supernatant was then removed with a 200 μ L pipet; and the remaining ethanol was allowed to evaporate at room temperature over 5 to 10 minutes. The remaining pellet was suspended in 2 mL water and purified by HPLC with a 50 mM aqueous triethylammonium acetate mobile phase at pH 7.5. The fractions containing the library were collected, pooled and lyophilized. The resulting residue was redissolved in 2.5 mL aqueous Na_2HPO_4 and 100 μ L piperidine was added, resulting in the formation of a precipitate. The precipitate was separated from the supernatant by centrifugation and washed with 200 μ L water. The wash and the supernatant were combined and used for Cycle 2.

Cycles 2-5

For each of these cycles, the combined solution resulting from the previous cycle was divided into 12 equal aliquots of 50 μ L each and placed in PCR tubes. To each tube was added a solution comprising a different tag, and ligation, purification and acylation were performed as described for Cycle 1, except that for Cycles 3-5, the HPLC purification step described for Cycle 1 was omitted. The correspondence between tags and building block precursors for Cycles 2-5 is presented in Table 2.

The products of Cycle 5 were ligated with the closing primer shown below, using the method described above for ligation of tags.

5' - PO_3 - GGCACATTGATTGGGAGTCA

10 GTGTAACTAAACCCCTCAGT - PO_3 - 5'

Table 2

Building Block Precursor	Cycle 2 Tag	Cycle 3 Tag	Cycle 4 Tag	Cycle 5 Tag
BB1	2.7	3.7	4.7	5.7
BB2	2.8	3.8	4.8	5.8
BB3	2.2	3.2	4.2	5.2
BB4	2.10	3.10	4.10	5.10
BB5	2.1	3.1	4.1	5.1
BB6	2.12	3.12	4.12	5.12
BB7	2.5	3.5	4.5	5.5
BB8	2.6	3.6	4.6	5.6
BB9	2.4	3.4	4.4	5.4
BB10	2.3	3.3	4.3	5.3
BB11	2.9	3.9	4.9	5.9
BB12	2.11	3.11	4.11	5.11

Results:

15 The synthetic procedure described above has the capability of producing a library comprising 12^5 (about 249,000) different structures. The synthesis of the library was monitored via gel electrophoresis of the product of each cycle. The gel electrophoresis shows that each cycle results in the expected molecular weight increase and that the products of each cycle are substantially homogeneous with regard to molecular weight.

Example 13.

20

Direct transfer acylation reaction.

This is an example of a stage 2 synthesis direct transfer reaction, involving the reactive group NH₂, and an activated ester, N-hydroxysuccinimide ester. Similar reaction conditions can be applied to the stage 1 acylation reaction, except that the concentration of the incoming molecule fragment must be higher (e.g., 100 mM incoming molecule fragment in a stage 1 synthesis). The example is taken from (Freskgård et al., WO 2004/039825 A2, example 3, p. 111-116).

The molecule fragment, in the following called "attachment entity (AE)" is in the following experiments either a scaffold molecule fragment, e.g. the peptide, CFFKKK, attached to an oligonucleotide, in the following called "Identifier", or a molecule fragment, in the following called "recipient reactive group" exemplified by an amino modified oligonucleotide. These molecule fragments allow transfer of three or one molecule fragments, respectively.

The identifier used in this experiment is an oligonucleotide coupled to the peptide CFFKKK as described in Example 4. The molecule fragment, in the following called "functional entity (FE)", is in this experiment 4-Pentynolic acid, which is attached to an oligonucleotide. The Identifier oligonucleotide, coupled to the CFFKKK scaffold, is annealed to the oligonucleotide carrying the 4-pentynolic acid, thereby bringing the two molecule fragments into close proximity. The annealing is directed by the complementarity of the two oligonucleotides.

The annealing was performed using 600 pmol of the 4-pentynolic acid oligonucleotide and 400 pmol Identifier oligonucleotide in 0.1 M MES buffer at 25°C in a shaker for 2 hours.

After annealing and subsequent reaction between the two molecule fragments, the sample was purified by micro-spin gel filtration and analyzed by MS. The observed mass was 7323.45 Da, which correspond well with the calculated mass, 7324.00 Da. Thus, the MS shows a mass corresponding to the transfer of the molecule fragment (4-pentenolic acid) onto the amino group of the Identifier oligonucleotide through formation of an amide bond.

Another example of transfer of a molecule fragment is shown below using the amine-modified oligonucleotide directly as the AE on the Identifier molecule. The functional entity on the building block molecule used in this experiment was 4-pentynolic acid.

The annealing was performed using 500 pmol of either carrier molecule in 0.1 M MES buffer and incubating the mixture at 25°C in a shaker for 2 hours. The molecule fragment (4-pentenolic acid) was transferred to the amino group on the Identifier molecule during the annealing (see below). After annealing and transfer the sample was purified by micro-spin gel filtration and analyzed by MS. The observed mass was 6398.04 Da, which correspond well with the calculated mass, 6400.00 Da. Thus, the MS spectra of the Identifier molecule after transfer of the functional entity show a mass corresponding to the transferred molecule fragment, 4-pentenolic acid, onto the Identifier molecule, by formation of an amide bond.

Another example of direct transfer of a molecule fragment by acylation uses the amine modified oligo directly as the Identifier molecule. The functional entity used in this experiment was Hexynolic acid. The annealing was performed using 500 pmol of either

carrier molecule in 0.1 M MES buffer incubated at 25 °C in a shaker for 2 hours. The hexynolic acid molecule fragment was transferred to the amino group on the Identifier molecule through formation of an amide bond (see below). After annealing and transfer the sample was purified by micro-spin gel filtration and analyzed by MS. The observed mass was 6411.96 Da, which correspond well with the 15 calculated mass, 6414 Da. Thus, the MS spectra show a mass corresponding to the transfer of hexynolic acid onto the amine of the Identifier oligo through amide bond formation.

Example 14.

10 Multi-Step stage 2 synthesis using different types of cleavable linkers.

This is an example of a multistep, stage 2 synthesis involving several carriers hybridizing to different positions of the same template, and the use of three different types of cleavable linkers, employed in indirect transfer reactions. Also described is a templated Wittig reaction, a direct transfer reaction. The description of the experiment is taken from (Liu et al., WO 2004/016767 A2, example 3, p. 112-117). The figures referred to are from the same patent application.

Three distinct strategies have been developed to link chemical reagents (reactive units) with their decoding DNA oligonucleotides, and to purify product after any DNA-templated synthetic step. When possible, an ideal reagent-oligonucleotide linker for DNA-templated synthesis positions the oligonucleotide as a leaving group of the reagent. Under this "autodeaving" linker strategy, the oligonucleotide-reagent bond is cleaved as a natural chemical consequence of the reaction (see WO 2004/016767 A2, Figure 28A).

As the first example of this approach applied to DNA-templated chemistry, a dansylated Wittig phosphorane reagent (WO 2004/016767 A2, compound (1)) was synthesized in which the decoding DNA oligonucleotide was attached to one of the aryl phosphine groups (Hughes (1996) TETRAHEDRON LETT. 37: 7595). DNA-templated Wittig olefination with aldehyde-linked template 2 resulted in the efficient transfer of the fluorescent dansyl group from the reagent to the template to provide olefin 3 (WO 2004/016767 A2, Figure 28A). As a second example of an autodeaving linker, DNA-linked thioester 4 (WO 2004/016767 A2), when activated with Ag(I) at pH 7.0 (Zhang et al. (1999) J. Am. Chem. Soc. 121: 3311) acylated amino-terminated template 5 to afford amide product 6 (WO 2004/016767 A2, Figure 28B).

Ribosomal protein biosynthesis uses aminoacylated tRNAs in a similar autodeaving linker format to mediate RNA-templated peptide bond formation. To purify desired products away from unreacted reagents and from cleaved oligonucleotides following DNA-templated reactions using autodeaving linkers, biotinylated reagent oligonucleotides and washing crude reactions with streptavidin-linked magnetic beads (see WO 2004/016767 A2, Figure 30A) were utilized. Although this approach

Desired products generated from DNA-templated reactions using the scarless, or useful scar linkers can be readily purified using biotinylated reagent oligonucleotides (WO 2004/016767 A2, Figure 30B). Reagent oligonucleotides together with desired products are first captured on streptavidin-linked magnetic beads. Any unreacted template bound to reagent by base pairing is removed by washing the beads with buffer containing 4 M guanidinium chloride. Biotinylated molecules remain bound to the streptavidin beads under these conditions. Desired product then is isolated in pure form by eluting the beads with linker cleavage buffer (in the examples above, either pH 11 or sodium periodate (NaIO₄)-containing buffer), while reacted and unreacted reagents remain bound to the beads.

As one example of a specific library generated as described above, three iterated cycles of DNA-templated amide formation, traceless linker cleavage, and purification with streptavidin-linked beads were used to generate a non-natural tripeptide (WO 2004/016767 A2, Figures 31A-B). Each 20 amino acid reagent was linked to a unique biotinylated 10-base DNA oligonucleotide through the sulfone linker described above. The 30-base amine-terminated template programmed to direct the tripeptide synthesis contained three consecutive 10-base regions that were complementary to the three reagents, mimicking the strategy that would be used in a multi-step DNA-templated small molecule library synthesis.

In the first step, two equivalents of **13** (see WO 2004/016767 A2) were activated by treatment with 20 mM EDC, 15 mM sulfo-NHS, 0.1 M MES buffer pH 5.5, and 1 M NaCl, for 10 minutes at 25 °C. The template then was added in 0.1 M MOPS pH 7.5, and 1M NaCl, at 25°C and was allowed to react for 1 hour. The free amine group in **14** (see WO 2004/016767 A2) then was elaborated in a second and third round of DNA-templated amide formation and linker cleavage to afford dipeptide **15** and tripeptide **16**

(see WO 2004/016767 A2) using the following conditions: two equivalents of reagent, 50 mM DMT-MM, 0.1 M MOPS buffer pH 7.0, 1 M NaCl, at 25 °C for 6 hours. Desired product after each step was purified by capture on avidin-linked beads and elution with 0.1 M CAPS buffer pH 11.8, 60 mM BME, at 37 °C for 2 hours. The progress of each reaction and purification was followed by denaturing polyacrylamide gel electrophoresis (WO 2004/016767 A2, Figure 31B, bottom). Lanes 3, 6, and 9 represent control reactions using reagents containing scrambled oligonucleotide sequences.

The progress of each reaction, purification, and sulfone linker cleavage step was followed by denaturing polyacrylamide gel electrophoresis. The final tripeptide linked to template **16** (see WO 2004/016767 A2) was digested with the restriction endonuclease *EcoBI* and the digestion fragment containing the tripeptide was characterized by MALDI mass spectrometry. Beginning with 2 nmol (~ 20 ug) of starting material, sufficient tripeptide product was generated to serve as the template for more than 10⁶ *in vitro* selections and PCR reactions (Kramer *et al.* (1999) CURRENT PROTOCOLS IN MOL. BIOL. 3: 15.1) (assuming 1/10,000 molecules survive selection). No significant product was generated when the starting material template was capped with acetic anhydride, or

does not separate reacted templates from unreacted templates, unreacted templates can be removed in subsequent DNA-templated reaction and purification steps.

Reagents bearing more than one functional group can be linked to their decoding DNA oligonucleotides through second and third linker strategies. In the "scarless linker" approach (WO 2004/016767 A2, Figure 28C), one functional group of the reagent is reserved for DNA-templated bond formation, while the second functional group is used to attach a linker that can be cleaved without introducing additional unwanted chemical functionality. The DNA-templated reaction then is followed by cleavage of the linker attached through the second functional group to afford desired products (WO 2004/016767 A2, Figure 28C). For example, a series of aminoacylation reagents such as (D)-Phe derivative **7** (WO 2004/016767 A2) were synthesized in which the alpha-amine is connected through a carbamoylthio/sulfone linker (Zarling *et al.* (1980) J. IMMUNOLOGY 124: 913) to its decoding DNA oligonucleotide. The product (WO 2004/016767 A2, compound (8)) of DNA-templated amide bond formation using this reagent and an amine-terminated template (WO 2004/016767 A2, (5)) was treated with aqueous base to effect the quantitative elimination and spontaneous decarboxylation of the linker, affording product **9** containing the cleanly transferred amino acid group (WO 2004/016767 A2, Figure 28C). This sulfone linker is stable in pH 7.5 or lower buffer at 25 °C for more than 24 hours yet undergoes quantitative cleavage when exposed to pH 11.8 buffer for 2 hours at 37 °C.

In some cases it may be advantageous to introduce one or more new chemical groups as a consequence of linker cleavage. Under a third linker strategy, linker cleavage generates a "useful scar" that can be functionalized in subsequent steps (WO 2004/016767 A2, Figure 28C). As an example of this class of linker, amino acid reagents such as the (L)-Phe derivative **10** were generated linked through 1,2-diols (Fruchart *et al.* (1999) TETRAHEDRON LETT. 40: 6225) to their decoding DNA oligonucleotides. Following DNA-templated amide bond formation with amine terminated template (WO 2004/016767 A2, compound (5)), this linker was quantitatively cleaved by oxidation with 50 mM aqueous sodium periodate (NaIO₄) at pH 5.0 to afford product **12** containing an aldehyde group appropriate for subsequent functionalization (for example, in a DNA-templated Wittig olefination, reductive amination, or nitroaldol addition).

Figure 29 of (WO 2004/016767 A2) shows the results of exemplary DNA-templated synthesis experiments using autodcleaving linkers, scarless linkers, and useful scar linkers. The depicted reactions were analyzed by denaturing PAGE. Lanes 1-3 were visualized using UV light without DNA staining; lanes 4-10 were visualized by staining with ethidium bromide following by UV-transillumination. Conditions for 1 to 3 were: one equivalent each of reagent and template, 0.1 M TAPS buffer pH 8.5, 1 M NaCl, at 25 °C for 1.5 hours. Conditions for 4 to 6 were: three equivalents of 4,0'-¹⁴ M. MES buffer pH 7.0., 1 M sodium nitrite (NaNO₂) 10 mM silver nitrate (AgNC<3>), at 37 °C for 8 hours. Conditions for 8 to 9 were 0.1 M 3-(cyclohexylamino)-1-5-propanesulfonic acid (CAPS) buffer pH 11.8, 60 mM (3-mercaptoethanol (BME), at 37 °C for 2 hours. Finally, conditions for 11 to 12 were: 50 mM aqueous NaIO₄, at 25 °C for 2 hours. RI = NH(CH₂)₂NH-dansyl; R₂ = biotin.

when control reagents containing sequence mismatches were used instead of the complementary reagents (WO 2004/016767 A2, Figure 31B).

A non-peptidic multi-step DNA-templated small molecule synthesis that uses all three linker strategies developed above was also performed (WO 2004/016767 A2, Figure 32A-32B). An

5 amine-terminated 30-base template was subjected to DNA-templated amide bond formation using an aminoacyl donor reagent (WO 2004/016767 A2, compound (17)) containing the diol linker and a biotinylated 10-base oligonucleotide to afford amide 18 (WO 2004/016767 A2) (two equivalents

17 in 20 mM EDC, 15 mM sulfo-NHS, 0.1 M MES buffer pH 5.5, 1 M NaCl, 10 minutes, 25 °C, then add to template in 0.1 M MOPS pH 7.5, 1M NaCl at 16°C for 8 hours). The desired product then was

10 isolated by capturing the crude reaction on streptavidin beads followed by cleaving the linker with NaIO₄ to generate aldehyde 19 (WO 2004/016767 A2). The DNA-templated Wittig reaction of 19 with the biotinylated autodeactivating phosphorane reagent 20 (WO 2004/016767 A2) afforded

fumaramide 21 (WO 2004/016767 A2) (three equivalents 20, 0.1 M TAPS pH 9.0, 3 M NaCl at 25 °C for 48 hours). The products from the second DNA-templated reaction were partially purified by

15 washing with streptavidin beads to remove reacted and unreacted reagent. In the third DNA-templated step, fumaramide 21 was subjected to a DNA-templated conjugate addition (Gartner et al. (2001) J. Am. Chem. Soc. 123: 6961) using thiol reagent 22 (WO 2004/016767 A2) linked

through the sulfone linker to a biotinylated oligonucleotide (three equivalents 22, 0.1 M TAPS pH 8.5, 1 M NaCl at 25°C for 21 hours). The desired conjugate addition product (WO 2004/016767 20 A2, compound (23)) was purified by immobilization with streptavidin beads. Linker cleavage with pH 11 buffer afforded final product 24 (WO 2004/016767 A2) in 5-10% overall isolated yield for the three bond forming reactions, two linker cleavage steps, and three purifications (WO 2004/016767 A2, Figures 32A-32B). The final product was digested with *EcoRI* and the mass of the small molecule-linked template fragment was confirmed by MALDI mass spectrometry (exact mass:

25 2568, observed mass: 2566±5). As in the tripeptide example, each of the three reagents used during this multi-step synthesis, annealed at a unique location on the DNA template, and control reactions with sequence mismatches yielded no product (WO 2004/016767 A2, Figure 32B, bottom). In Figure 32B, bottom lanes 3, 6, and 9 represent control reactions. As expected, control reactions in which the Wittig reagent was omitted (step 2) also did not generate product following the third step. Taken together, the DNA-templated syntheses of compounds 16 and 24 (see WO 30 2004/016767 A2) demonstrate the ability of DNA to direct the sequence-programmed multi-step synthesis of both oligomeric and non-oligomeric small molecules: unrelated in structure to nucleic acids.

35 Example 15:

Stage 2 reactions in organic solvents

This is an example of a stage 2 synthesis performed in organic solvents. Similar or identical conditions can be applied to stage 1 synthesis, except that the concentrations of molecule fragments must be appropriately high to obtain efficient reaction, e.g. higher concentrations of molecule fragments than 10 mM. The description of the experiment is taken from (Liu et al., WO 5 2004/016767 A2, example 4, p. 117-118). The figures referred to are from the same patent application.

A variety of DNA-templated reactions can occur in aqueous media. It has also been discovered that DNA-templated reactions can occur in organic solvents, thus greatly expanding the scope of DNA-templated synthesis. Specifically, DNA templates and reagents have been complexed with long chain tetraalkylammonium cations (see, Jost et al. (1989) NUCLEIC ACIDS RES. 17:2143; Melnikov et al. (1999) LANGMUIR 15: 1923-1928) to permit quantitative dissolution of reaction components in anhydrous organic solvents including CH₂Cl₂, CHCl₃, DMF and methanol. Surprisingly, it was found that DNA-templated synthesis can indeed occur in anhydrous organic solvents with high sequence selectivity.

15 Figure 33, WO 2004/016767 A2 shows DNA-templated amide bond formation reactions where the reagents and templates are complexed with dimethyldodecylammonium cations either in separate vessels or after preannealing in water, lyophilized to dryness, dissolved in CH₂Cl₂, and mixed together. Matched, but not mismatched, reactions provided products both when reactants were preannealed in aqueous solution and when they were mixed for the first time in CH₂Cl₂ (WO 20 2004/016767 A2, Figure 33). DNA-templated amide formation and Pd-mediated Heck coupling in anhydrous DMF also proceeded sequence-specifically.

These observations of sequence-specific DNA-templated synthesis in organic solvents imply the presence of at least some secondary structure within tetraalkylammonium-complexed DNA in organic media, and should permit DNA receptors and catalysts to be evolved towards stereoselective binding or catalytic properties in organic solvents. Specifically, DNA-templated reactions that are known to occur in aqueous media, including conjugate additions, cycloadditions, displacement reactions, and Pd-mediated couplings can also be performed in organic solvents.

It is contemplated that reactions in organic solvents may be utilized that are inefficient or impossible to perform in water. For example, while Ru-catalyzed olefin metathesis in water has been reported (Lynn et al. (1998) J. AM. CHEM. SOC. 120: 1627-1628; Lynn et al. (2000) J. AM. CHEM. SOC. 122: 6601-6609; Mohr et al. (1996) ORGANOMETALLICS 15: 4317-

4325), the aqueous metathesis system is extremely sensitive to the identities of the functional groups. The functional group tolerance of Ru-catalyzed olefin metathesis in organic solvents, however, is significantly more robust. Some exemplary reactions to utilize in organic solvents include, but are not limited to, 1,3-dipolar cycloaddition between nitrones and olefins which can proceed through transition states that are less polar than ground state starting materials.

Example 16:**Stage 2 Omega synthesis (subprocess F), involving amine acylation, Wittig olefination, 1,3-Dipolar Cycloaddition and Reductive amination.**

5 This is an example of a stage 2 synthesis employing the Omega DNA architecture during the templated synthesis. Also described are the conditions allowing amine acylation, Wittig olefination, 1,3-Dipolar Cycloaddition and Reductive amination reactions to proceed efficiently. The same conditions can be applied during a stage 1 synthesis involving the same reactions, except that the molecule fragments must be added at a higher concentration (e.g. 10-100 mM molecule fragment). The description of the experiments is taken from (Liu et al., WO 2004/016767 A2, example 5, p. 118-126). The figures referred to are from the same patent application.

This example discloses two different template architectures that further expand the scope of nucleic acid-templated synthesis. During a nucleic acid-templated chemical reaction a portion of a template anneals to a complementary sequence of an oligonucleotide-linked reagent, holding functional groups on the template and transfer unit in reactive proximity. Template architecture can have a profound effect on the nature of the resulting reaction, raising the possibility of manipulating reaction conditions by rationally designing template-reagent complexes with different secondary structures. It was hypothesized that the distance dependence of certain DNA-templated reactions such as 1,3-dipolar cycloadditions and reductive animation could be overcome by designing a new architecture that permits a reagent to anneal to two distinct and spatially separated regions of the template. In the "Omega" architecture (see WO 2004/016767 A2, Figure 7), the template oligonucleotide contains a small number of constant bases at, for example, the reactive 5'-end of the template in addition to distal coding regions. The oligonucleotide of the transfer unit for the Omega architecture contains at its reactive 3'-end the bases that complement the constant region of the template followed by bases that complement a coding region anywhere on the template. The constant regions were designed to be of insufficient length to anneal in the absence of a complementary coding region. When the coding region of the template and transfer unit are complementary and anneal, the elevated effective molarity of the constant regions induces their annealing. Constant region annealing forms a bulge in the otherwise double-stranded template-reagent complex and places groups at the ends of the template and reagent in reactive proximity. This design permits distance-dependent DNA-templated reactions to be encoded by bases distal from the reactive end of the template.

The efficiency of DNA-templated synthesis using the Omega architecture was compared with that of the standard E and H architectures. The Omega architectures studied comprise (i) three to five constant bases at the 5' end of the template followed by (ii) a five- to 17-base loop and (iii) a ten-

base coding region. As a basis for comparison, four different classes of DNA-templated reactions were performed that collectively span the range of distance dependence observed to date.

Amine acylation reactions are representative of distance independent reactions that proceed efficiently even when considerable distances (e.g., 30 bases) separate the amine and carboxylate groups. As expected, amine acylation (20 mM DMT-MM, pH 7.0, at 30 °C for 12 hours) proceeded efficiently (46-96% yield) in all architectures with both small and large distances between reactive groups on the reagent and template (WO 2004/016767 A2, Figure 34, lanes 1-5; and Figure 35A).

The Omega architecture mediated efficient amine acylation with three, four, or five constant bases at the reactive ends of the template and reagent and 10 or 20 bases between annealed reactants ($n = 10$ or 20). Importantly, control reactions in which the distal coding region contained three sequence mismatches failed to generate significant product despite the presence of the complementary three- to five-base constant regions at the ends of the template and reagent 5 (see WO 2004/016767 A2, Figure 34, lane 5 for a representative example). The Omega architecture, therefore, did not impede the efficiency or sequence-specificity of the distance-independent amine acylation reaction.

DNA-templated Wittig olefination reactions proceed at a significantly lower rate when the aldehyde and phosphorane are separated by larger numbers of template bases, even though product yields typically are excellent after 12 hours or more of reaction regardless of intervening distance. After only 2 hours of reaction (pH 7.5, 30 °C) in the E or H architectures, however, yields of olefin products were three- to six-fold lower when reactants were separated by ten or more bases ($n = 10$ or 20) than when reactants are separated by only one base ($n = 1$) (WO 2004/016767 A2, Figure 34, lanes 6-7, and Figure 35B). In contrast, the Omega architecture with four or five constant bases at the reactive end resulted in efficient and sequence-specific Wittig product formation after 2 hours of reaction even when 10 or 20 bases separated the coding region and reactive end of the template (WO 2004/016767 A2, Figure 34, lanes 8-9, and Figure 35B). These results suggest that the constant regions at the reactive ends of the template and transfer unit in the Omega architecture permit the aldehyde and phosphorane moieties to react at an effective concentration comparable to that achieved with the E-architecture when $n = 1$ (WO 2004/016767 A2, Figure 34).

Among the many DNA-templated reactions studied to date, the 1,3-dipolar cycloaddition and reductive animation reactions demonstrate the most pronounced distance dependence. Both reactions proceed in low to modest efficiency (7%-44% yield) under standard reaction conditions using the E or H architectures when 10 or 20 bases separate the annealed reactive groups (WO 2004/016767 A2, Figure 34, lanes 10-11 and 14-15, and Figures 35C-35D). This distance dependence limits the positions on a DNA template that can encode these or other similarly distant dependent reactions. In contrast, both 1,3-dipolar cycloaddition and reductive animation proceed efficiently (up to 97% yield) and sequence-specifically when encoded by template bases 15-25 bases away from the functionalized end of the template using the Omega architecture with four or

five constant bases (WO 2004/016767 A2, Figure 34, lanes 12-13 and 16-17, and Figures 35C-35D). These results demonstrate that the templates Omega architecture permits distance-dependent reactions to be efficiently directed by DNA bases far from the reactive end of the template. By overcoming the distance dependence of these reactions while preserving the efficiency of distant independent reactions, the Omega architecture may permit virtually any contiguous subset of bases in a single-stranded 30-base template to encode any viable DNA-templated reaction. Interestingly, the Omega templates with only three constant bases at their reactive ends do not consistently improve the efficiency of these reactions compared with the E-architecture (WO 2004/016767 A2, Figures 35C-35D), suggesting that four or five constant bases may be required in the Omega architecture to fully realize favorable proximity effects.

In order to probe the structural features underlying the observed properties of the Omega architecture, the thermal denaturation of the Omega-5 and E architectures using $n = 10$ and $n = 20$ reagents were characterized. For all template-reagent combinations, only a single cooperative melting transition was observed. Compared to the E architecture reagent lacking the five-base constant region, the Omega-5 reagent increased the hypochromicity upon annealing by ~50% but did not significantly affect melting temperature in either phosphate-buffered saline (PBS) or in 50 mM sodium phosphate pH 7.2 with 1 M NaCl (WO 2004/016767 A2, Figure 36). These results are consistent with a model in which template-reagent annealing in the Omega architecture is dominated by coding region interactions even though the constant region forms secondary structure once the coding region is annealed. The entropic cost of partially ordering the loop between the coding and constant regions may, therefore, be offset by the favorable interactions that arise upon annealing of the constant region.

DNA templates of arbitrary length are easy to synthesize and undesired cross-reactivity between reactants in the same solution can be avoided using concentrations that are too low to allow non-complementary reactants to react intermolecularly. These features of DNA-templated synthesis permit more than one DNA-templated reaction to take place on a single template in one solution, saving the effort associated with additional DNA-templated steps and product purifications. Multiple DNA-templated reactions per step can be difficult using the E, H, or Omega architectures, because the reagent oligonucleotide that remains annealed to the template following the first reaction forms a relatively rigid double helix that can prevent a second reagent annealing further away along the template from encountering the reactive end of the template. To overcome this, the reactive group on the template was moved from the end of the oligonucleotide to the middle, attaching the reactive group to the non-Watson-Crick face of a base. This "T" architecture (see WO 2004/016767 A2, Figure 7G) was designed to permit two DNA-templated reactions, one with a reagent coupled to the 5' end of the oligonucleotide of a first transfer unit and one with a reagent coupled to the 3'-

end of the oligonucleotide of a second transfer unit, to take place sequence-specifically in the same solution on a single template.

To test the viability of the T-architecture in DNA-templated reactions, the efficiency of the amine acylation, Wittig olefination, 1,3-dipolar cycloaddition, and reductive amination reactions using the S architecture was studied. The T architecture sequence-specifically directed these four reactions with efficiencies comparable to or greater than those of the E or H architectures (WO 2004/016767 A2, Figure 37, 69-100% yield when $n = 1$).

It can thus be concluded that it is possible to perform each of those reactions in an efficient way, providing high yields, at least for one DNA architecture.

10

The observed degree of distance dependence using the T architecture for each of the four reactions was consistent with the above findings (WO 2004/016767 A2, compare Figure 37 and Figure 35). Together these results demonstrate that the T architecture can mediate sequence-specific and efficient DNA-templated synthesis.

Once the ability of the T architecture to support efficient DNA-templated synthesis was established, the ability of the T architecture to direct two DNA-templated reactions on one template in one solution was studied. Two different two-reaction schemes using the T architecture were performed. In the first scheme, depicted in (WO 2004/016767 A2, Figure 38A), a benzaldehyde-linked T template (WO 2004/016767 A2, (1)) was combined with a phosphine-linked reagent (WO

2004/016767 A2, (2)) and an alpha-iodoamide-linked reagent (WO 2004/016767 A2, (3)) in a single solution (pH 8.5, 1 M NaCl, at 25 °C for 1 hour). The phosphine-linked oligonucleotide complemented ten bases of the template 5' of the aldehyde ($n = -4$), while the iodide-linked oligonucleotide complemented ten bases 3' of the aldehyde ($n = 0$). DNA-templated Sn2 reaction between the phosphine and alpha-iodoamide generated the corresponding phosphorane, which then participated in a DNA-templated Wittig reaction to generate cinnamaldehyde 4 (WO 2004/016767 A2) in 52% overall yield after 1 hour (Figure 38B, lanes 9-10). Control reactions containing sequence mismatches in either reagent generated no detectable product. The additional control reaction lacking the aldehyde group on the template generated only the Sn2 reaction product (Figure 38B, lanes 3-4) while control reactions lacking either the phosphine group or the alpha-iodoamide group did not generate any detectable products (Figure 38B, lanes 5-8).

In a second two-reaction scheme mediated by the T architecture, depicted in (WO 2004/016767 A2, Figure 38C), an amine-linked T template (WO 2004/016767 A2, (5)) was combined with a propargylglycine-linked 5' reagent (WO 2004/016767 A2, (6)) at $n = -1$ and a phenyl azide-linked 3' reagent (WO 2004/016767 A2, (7)) at $n = 1$. The addition of 20 mM DMT-MM at pH 7.0 to induce amide formation followed by the addition of 500 uM copper(II) sulfate and sodium ascorbate to induce the recently reported Sharpless-modified Huisgen 1,3-dipolar cycloaddition provided 1,4-disubstituted triazole alanine adduct 8 (WO 2004/016767 A2) in 32% overall yield.

reagent (WO 2004/016767 A2, (6)) at $n = -1$ and a phenyl azide-linked 3' reagent (WO 2004/016767 A2, (7)) at $n = 1$. The addition of 20 mM DMT-MM at pH 7.0 to induce amide formation followed by the addition of 500 uM copper(II) sulfate and sodium ascorbate to induce the recently reported Sharpless-modified Huisgen 1,3-dipolar cycloaddition provided 1,4-disubstituted triazole alanine adduct 8 (WO 2004/016767 A2) in 32% overall yield.

1,4-disubstituted triazole alanine adduct 8 (WO 2004/016767 A2) in 32% overall yield.

Taken together, these observations show that the T architecture permits two sequence-specific DNA-templated reactions to take place on one template in one solution.

Importantly, the T architecture templates described above were accepted as efficient templates for both a single cycle of primer extension as well as standard PCR amplification using Taq DNA polymerase, consistent with the known tolerance of several DNA

polymerases for modifications to the non-Watson-Crick face of DNA templates. In addition to reducing the number of separate DNA-templated steps needed to synthesize a target structure, this architecture may also permit three-component reactions commonly used to build structural complexity in synthetic libraries to be performed in a DNA-templated format.

In summary, the Omega and T architectures significantly expand the scope of DNA-templated synthesis. By enabling distance-dependent DNA-templated reactions to be encoded by bases far away from the reactive end of the template, the Omega architecture expands the types of reactions that can be encoded anywhere on a DNA template. The T architecture permits two DNA-templated reactions to take place on a single template in one step.

Materials and Methods

Oligonucleotide synthesis. Unless otherwise specified, DNA oligonucleotides were synthesized and functionalized as previously described using 2'-[2-(4-monomethoxytrityl)]

20 aminoethoxy]ethyl-(2-cyanoethyl)-N,N-dilisopropyl-phosphoramidite (Glen Research, Sterling, Virginia, USA) for 5'-functionalized oligonucleotides, and using (2'-dimethoxytrityl)oxymethyl-6-fluorenylmethoxycarbonylamino-hexane-I-succinoyl]-long chain alkylamino-CPG (Glen Research, Sterling, Virginia, USA) for 3'-functionalized oligonucleotides (Calderone et al. (2002) ANGEW. CHEM. INT. ED. ENGL. 41:4104; (2002) ANGEW. CHEM. 114: 4278). In the case of templates for the T architecture, amine groups were added using 5'-dimethoxytrityl-5-[N-(trifluoroacetylaminohexyl)-3-acrylimido]-2'-deoxyuridine-3'-[(2-cyanoethyl)-(N,N-dilisopropyl)]-phosphoramidite (Glen Research, Sterling, Virginia, USA) and then acylated as reported previously (Calderone et al. (2002) *supra*).

30 **Amine Acylation.** Amine-labeled and carboxylic acid-labeled DNA were combined in aqueous 100 mM MOPS buffer, 1 M NaCl, pH 7.0 (60 nM in template DNA, 120 nM in reagent DNA) in the presence of 20 mM DMT-MM. Reactions proceeded for 12 hours at 25 °C.

35 **Wittig Olefin ation.** Aldehyde-labeled and phosphorane-labeled DNA were combined in aqueous 100 mM MOPS, 1 M NaCl, pH 7.5 (60 nM in template DNA, 120 nM in reagent DNA). Reactions proceeded for 2 hours at 30°C.

1,3-Dipolar Cycloaddition. Dialdehyde-labeled DNA was incubated in 260 mM

N-methylhydroxylamine hydrochloride for 1 hour at room temperature (Gartner et al. (2002) J. AM. CHEM. SOC. 124: 10304). It was subsequently combined with succinimide-labeled DNA in aqueous 50 mM MOPS, 2.8 M NaCl, pH 7.5 (final concentrations of N-methylhydroxylamine hydrochloride 0.75 mM, 60 nM in template DNA and 9.0 nM in reagent DNA). Reactions proceeded for 12 hours at 37°C.

Reductive Amination. Amine-labeled and aldehyde-labeled DNA were combined in aqueous 100 mM MES buffer, 1 M NaCl, pH 6.0 (60 nM in template DNA, 120 nM in reagent DNA). Sodium cyanoborohydride was added as a 5 M stock in 1 M NaOH to a final concentration of 38 mM, and reactions proceeded for 2 hours at 25 °C. Reactions were quenched by 10 ethanol precipitation in the presence of 15 mM methylamine.

T Architecture-mediated Conversion of Compound 1 to 4. The 5'-phosphine-linked oligonucleotide (WO 2004/016767 A2, (2)) was generated by coupling N-succinimidylodacetate (SIA) to the amine derived from 12-(4-monomethoxytritylamino) dodecyl-(2-cyanoethyl)-(N,N-dilisopropyl)-phosphoramidite (Glen Research, Sterling, Virginia, USA) using the T (n = -4) oligonucleotide listed below, followed by treatment with 4-diphenylphosphinobenzoic acid as described previously (Gartner et al. (2002) *supra*).

The 3'-omega-iodoamide-linked reagent (WO 2004/016767 A2, (3)) was prepared by reacting the T (n = 1) oligonucleotide (see below) with SIA as described previously (Gartner et al. (2001) *supra*). Aldehyde-labeled template (WO 2004/016767 A2, (1)) was prepared by reacting the "T template" oligonucleotide (see below) with para-formyl benzoic acid N-hydroxysuccinimide ester as described previously (Gartner et al. (2002) ANGEW. CHEM. INT. ED. 41: 1796; (2002) ANGEW. CHEM. 114:1874). Template 1 was combined with reagents 2 and 3 (WO 2004/016767 A2) in aqueous 200 mM N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES) buffer at pH 8.5 with 1 M NaCl, (63 nM template and 125 nM of each reagent). Reactions proceeded for up to 1 hour at 25 °C.

The results of denaturing polyacrylamide gel electrophoresis analysis of these reactions is shown in (WO 2004/016767 A2, Figure 38B). The 30-base T architecture template (WO 2004/016767 A2, (1)) containing an aldehyde group was present in lanes 1-2 and lanes 5-10. A template lacking the aldehyde group but otherwise identical to (1) was present in lanes 3 and 4. DNA-linked phosphine reagent (WO 2004/016767 A2, (2)) was present in lanes 3-6 and lanes 9-10. DNA-linked alpha-iodoamide reagent (WO 2004/016767 A2, (3)) was present in lanes 3-4 and lanes 7-10. Lanes 1, 3, 5, 7 and 9 show reactions after 30 minutes. Lanes 2, 4, 6, 8, and 10 show reactions after 1 hour.

T Architecture-mediated Conversion of Compound 5 to 8. The 5'-propargylglycine linked oligonucleotide (WO 2004/016767 A2, (6)) was generated by combining the corresponding T (n = -1) 5'-amine-linked reagent oligonucleotide (see below) with 2 mg/mL

bis(sulfosuccinimidyl)suberate in 9:1 200 mM sodium phosphate pH 7.2:DMF for 10 minutes at 25 °C, followed by treatment with 0.3 vol of 300 mM racemic propargylglycine in 300 mM NaOH for 2

hours at 25 °C. The 3'-azido linked oligonucleotide (WO 2004/016767 A2, (7)) was generated by combining the T ($n = 1$) amine-linked reagent oligonucleotide (see below) with 2 mg/mL (N-15 hydroxysuccinimide)-4-azidobenzoate in 9:1 200 mM sodium phosphate pH 7.2:DMF for, 2 hours at 25 °C. Reagents 6 and 7 (WO 2004/016767 A2) were purified by gel filtration and reverse-phase HPLC. Template 5 and reagents 6 and 7 were combined in aqueous 100 mM MOPS pH 7.0 in the presence of 1 M NaCl and 20 mM DMT-MM for 12 hours (60 nM template, 120 nM reagents) at 25 °C. Copper (II) sulfate pentahydrate and sodium ascorbate were then added to 500 μ M each. After 1 hour at 25 °C, reactions were quenched by ethanol precipitation.

DNA Oligonucleotide Sequences Used. E or Omega template: 5'-H₂N-GGT ACGAAT TCG ACT

10 CGG GAA TAC CAC CTT [SEQ ID NO: 58]. H template: 5'-H₂N-CGC GAG CGT ACG CTC GCG GGT ACG AAT TCG ACT CGG GAA TAC CAC CTT [SEQ ID NO: 59]. T template: 5'-GGT ACG AAT TCG AC(TT-NH₂) CGG GAA TAC CAC CTT [SEQ ID NO: 60]. E or H reagent ($n = 1$): 5'-AAT TCG TAC C-NH₂ [SEQ ID NO: 61]. E or H reagent ($n = 10$): 5'-TCC CGA GTC G-NH₂ [SEQ ID NO: 62]. E or H reagent ($n = 20$): 5'-AAG GTG GTA T-NH₂ [SEQ ID NO: 63]. Mismatched E or H reagent: 5'-TCC CTG ATC G-NH₂ [SEQ ID NO: 64]. Omega-3 reagent ($n = 10$): 5'-TCC CGA GTC C-NH₂ [SEQ ID NO: 65]. Omega-4 reagent ($n = 10$): 5'-TCC CGA GTC C-NH₂ [SEQ ID NO: 66]. Omega-5 reagent ($n = 10$): 5'-TCC CGA GTC ACC-NH₂ [SEQ ID NO: 67]. Omega-3 reagent ($n = 20$): 5'-AAG GTG GTA TAC C-NH₂ [SEQ ID NO: 68]. Omega-4 reagent ($n = 20$): 5'-AAG GTG GTA TGT ACC-NH₂ [SEQ ID NO: 69]. Omega-5 reagent ($n = 20$): 5'-AAG GTG GTA TGT ACC-NH₂ [SEQ ID NO: 70]. Mismatched Omega-3 reagent: 5'-TCC CTG ATC GAC C-NH₂ [SEQ ID NO: 71]. Mismatched Omega-4 reagent: 5'-TCC CTG ATC GTA C-NH₂ [SEQ ID NO: 72].

Mismatched Omega-5 reagent: 5'-TCC CTG ATC GGT ACC-NH₂ [SEQ ID NO: 73]. T reagent ($n = 1$): 5'-GGT '5 ATT CCC G-NH₂ [SEQ ID NO: 74]. T reagent ($n = 2$): 5'-TGG TAT TCC C-NH₂ [SEQ ID NO: 75]. T reagent ($n = 3$): 5'-GTG GTA TTC C-NH₂ [SEQ ID NO: 76]. T reagent, ($n = 4$): 5'-GGT GGT ATT C-NH₂ [SEQ ID NO: 77]. T reagent ($n = 5$): 5'-AGG TGG TAT T-NH₂ [SEQ ID NO: 78]. T reagent ($n = 1$): 5'-NH₂-GTC GAA TTC G [SEQ ID NO: 79]. T reagent ($n = 4$) for 2: 5'-[C₁₂-amine linker]-AAT TCG TAG C [SEQ ID NO: 80]. Reaction yields were quantitated by denaturing polyacrylamide gel electrophoresis followed by ethidium bromide staining, UV visualization, and CCD-based densitometry of product and template starting material bands. Yield calculations assumed that templates and products were denatured and, therefore, stained with comparable intensity per base; for those cases in which products are partially double-stranded during quantitation, changes in staining intensity may result in higher apparent yields. Representative reaction products were characterized by MALDI mass spectrometry in addition to denaturing polyacrylamide 35 gel electrophoresis.

Melting curves were obtained on a Hewlett-Packard 8453 UV-visible spectrophotometer using a Hewlett-Packard 89090A Peltier thermocontroller. Absorbances of template-reagent pairs (1.5 μ M each) at 260 nm were measured every 1 °C from 20 °C to 80 °C holding for 1 minute at each temperature in either phosphate-buffered saline ("PBS," 137

mM NaCl, 2.7 mM potassium chloride, 1.4 mM potassium phosphate, 10 mM sodium phosphate, pH 7.4) or in high salt phosphate buffer ("HSB," 50 mM sodium phosphate pH 7.2, 1 M NaCl).

5

Example 18.

Functionalisation of oligonucleotides.

This is an example of how oligonucleotides may be functionalized for their further manipulation in stage 1 or stage 2 synthesis schemes. It also describes a stage 1 amine acylation reaction. The description of the experiment is taken from (Liu et al., WO 2004/016767 A2, p. 131). The figures referred to are from the same patent application.

2-bromopropionamide-NHS esters. 200 mg JV-hydroxysuccinimide (Pierce, Rockford, IL, USA) was dissolved in anhydrous CH₂Cl₂ together with 1.1 equivalents of a 2-bromopropionic acid (either racemic, (*R*)-, or (*S*)-) and 2 equivalents of *t*-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC) (Aldrich). The 2-bromopropionic acid enantiomers were >95% enantiopure as judged by chiral HPLC (5% isopropanol in hexanes, (*R*,*R*) WHELK 01 chiral phase, detection at 220 nm). The reaction was maintained at room temperature and complete after 1.5 hours as judged by TLC (EtOAc). The crude reaction mixture was extracted with 2.5% sodium hydrogen sulfate (NaHSO₄) to remove the excess EDC. The organic phase was washed with brine, dried over magnesium sulfate (MgSO₄), and concentrated *in vacuo*. The residue was dried and used directly for DNA functionalization.

5'-functionalization of oligonucleotides. An NHS ester prepared as described above was dissolved in DMSO. Up to 150 μ g of a 5'-amino DNA oligonucleotide was combined with 3 mg/mL NHS ester (final reaction = 10% DMSO) in 200 mM sodium phosphate (pH = 7.2) at room temperature for 2 hours. The functionalized oligonucleotides were purified by gel filtration and reverse-phase HPLC, and were characterized by denaturing PAGE and MALDI-TOF mass spectrometry.

30 3'-thiol modified oligonucleotides. The 3' thiol group was incorporated by standard automated DNA synthesis using 3'-disulfide-linked CPG (Glen Research, Sterling, 20 Virginia, USA). Following oligonucleotide synthesis, the disulfide was cleaved with 50 mM DTT, 1M TAPS (pH = 8.0) at room temperature for 1 hour and purified by gel filtration before being used in DNA-templated reactions.

35

Example 19.

One-pot simultaneous Stage 2 synthesis involving amine conjugate addition, thiol conjugate addition, nitro-Michael addition, reductive amination, amine acylation, and Wittig olefination.

5 This is an example of a number of templated reactions that are executed simultaneously in one solution, giving high yields of all reaction types tested. The same reaction conditions may be applied to stage 1 synthesis, except that the molecule fragments must be added in higher concentrations (preferably 10-100 mM). The description of the experiments is taken from (Liu et al., WO 2004/016767 A2, example 7, p. 137-142). The figures referred to are from the

10 same patent application.

This example demonstrates that oligonucleotides can simultaneously direct several different synthetic reaction types within the same solution, even though the reactants involved would be cross-reactive and, therefore, incompatible under traditional synthesis conditions. These findings also demonstrate that it is possible to perform a one-pot diversification of

15 synthetic library precursors into products using multiple, simultaneous and not necessarily compatible reaction types.

The ability of DNA templates to mediate diversification using different reaction types without spatial separation was initially tested by preparing three oligonucleotide templates of different DNA sequences (**1a-3a**) (WO 2004/016767 A2) functionalized at their 5' ends

20 with maleimide groups and three oligonucleotide reagents (**4a-6a**) (WO 2004/016767 A2) functionalized at their 3' ends with an amine, thiol, or nitroalkane group, respectively (WO 2004/016767 A2, Figure 46). The DNA sequences of the three reagents each contained a different 10-base annealing region that was complementary to ten bases, near the 5' end of each of the templates. Combining **1a** with **4a**, **2a** with **5a**, or **3a** with **6a** in three separate

25 vessels at pH 8.0 resulted in the expected DNA-templated amine conjugate addition, thiol conjugate addition, or nitro-Michael addition products **7-9** (WO 2004/016767 A2, Figure 46, lanes 1-3).

To distinguish the nine possible reaction products that could be generated upon combining **1a-6a**, the lengths of template oligonucleotides were varied to include 11, 17, or 23 bases

30 and the lengths of reagent oligonucleotides were varied to include 14, 16, or 18 bases. Differences in oligonucleotide length were achieved using extensions distal from the reactive groups that did not significantly affect the efficiency of DNA-templated reactions. This design permitted all nine possible reaction products (linked to 25, 27, 29, 31, 33, 35, 37, 39, or 41 bases of DNA) to be distinguished by denaturing polyacrylamide gel

35 electrophoresis.

A solution containing all three templates (**1a-3a**) was combined with a solution containing all three reagents (**4a-6a**) at pH 8.0. The resulting reaction exclusively generated the three desired products 7, 8, and 9 of lengths 25, 33, and 41 bases indicating that only the

three reactions corresponding to the complementary template-reagent pairs took place (WO 2004/016767 A2, Figure 46, lane 4). Formation of the other six possible reaction products was not detected by densitometry (<5% reaction). In contrast, individually reacting templates and reagents containing the same, rather than different, 10-base

5 annealing regions permitted the formation of all possible products (WO 2004/016767 A2, Figure 46, lane 5). This result demonstrates the ability of DNA-templated synthesis to direct the selective one-pot transformation of a single functional group into three distinct types of products (in this example, maleimide into secondary amine, thioether, or a-branched nitroalkane).

10 To test the ability of this diversification mode to support one-pot reactions requiring non-DNA-linked accessory reagents, an analogous experiment was conducted with two aldehyde-linked reagents either 14 or 16 bases in length (WO 2004/016767 A2, (4b) or (5b), respectively) and a complementary 11-base amine-linked template (WO 2004/016767 A2, (1b)) or a 17-base phosphorane-linked template (WO 2004/016767 A2, (2b)). Combining

15 1b and 4b at pH 8.0 in the presence of 3 mM NaBH₃CN resulted in the DNA-templated reductive amination product **10** (WO 2004/016767 A2), while 2b and 5b under the same conditions generated Wittig olefination product **11** (WO 2004/016767 A2, Figure 46). Mixing all four reactants together in one pot resulted in an identical product distribution as the combined individual Wittig olefination or reductive amination reactions (WO 2004/016767

20 A2, Figure 46). No reaction between amine 1b and aldehyde 5b or between phosphorane 2b and aldehyde 4b was detected (WO 2004/016767 A2, Figure 46, lane 8 versus lane 9).

The generality of this approach was explored by including multiple reaction types that required different accessory reagents. Three amine-linked templates (**1c-3c**) (WO 2004/016767 A2) of length 11, 17, or 23 bases were combined with an aldehyde-,

25 carboxylic acid-, or maleimide-linked reagent (4c-6c) (WO 2004/016767 A2) 14, 16, or 18 bases in length, respectively, at pH 8.0 in the presence of 3 mM NaBH₃CN, 10 mM 1-(3-dimethyl-aminopropyl)-3-ethylcarbodiimide (EDC), and 7.5 mM *N*-hydroxysulfosuccinimide (sulfo-NHS). The reactions containing all six reactants afforded the same three reductive

30 animation, amine acylation, or conjugate addition products (**12-14**) (WO 2004/016767 A2) that were generated from the individual reactions containing one template and one reagent and did not produce detectable quantities of the six possible undesired products arising from non-DNA-templated reactions (WO 2004/016767 A2, Figure 46, lanes 10-14).

Collectively, these results indicate that DNA-templated synthesis can direct simultaneous reactions between several mutually cross-reactive groups in a single pot to yield only the

35 sequence-programmed subset of many possible products.

The above three examples each diversified a single functional group (maleimide, aldehyde, or amine) into products of different reaction types. A more general format for the one-pot diversification of a DNA-templated synthetic library into products of multiple reaction types

would involve the simultaneous reaction of different functional groups linked to both reagents and templates. To examine this possibility, six DNA-linked nucleophile templates (15-20) (WO 2004/016767 A2) and six DNA-linked electrophile reagents (21-25) (WO 2004/016767 A2) collectively encompassing all of the functional groups used in the above

- 5 three examples (amine, aldehyde, maleimide, carboxylic acid, nitroalkane, phosphorane, and thiol) were prepared (WO 2004/016767 A2, Figure 47). These twelve DNA-linked reactants could, in theory, undergo simultaneous amine conjugate addition, thiol, conjugate addition, nitro-Michael addition, reductive amination, amine acylation, and Wittig olefination in the same pot, although the apparent second order rate constants of these six reactions vary by more than 10-fold.

Determining the outcome of combining all twelve reagents and templates in a single pot by using oligonucleotides of varying lengths; is difficult due to the large number (at least 28) of possible products that could be generated. Accordingly, the length of the reagents as 15, 20, 25, 30, 35, or 40 bases were varied but the length of the templates was fixed at 11 bases (WO 2004/016767 A2, Figure 47). Each of the six complementary template-reagent pairs when reacted separately at pH 8.0 in the presence of 3 mM NaBH₃CN; 10 mM EDC, and 7.5 mM sulfo-NHS generated the expected amine conjugate addition, thiol conjugate addition, nitro-Michael addition, reductive amination, amine acylation, or Wittig olefination products (WO 2004/016767 A2, Figure 47). Reaction efficiencies were greater than 50% relative to the corresponding individual reactions despite having to compromise between differing optimal reaction conditions. Templates 15-20 (WO 2004/016767 A2) were also prepared in a 3'-biotinylated form. The biotinylated templates demonstrated reactivities indistinguishable from those of their non-biotinylated counterparts (WO 2004/016767 A2, Figure 47).

- Six separate reactions each containing twelve reactants then were performed at pH 8.0 in the presence of 3 mM NaBH₃CN, 10 mM EDC, and 7.5 mM sulfo-NHS (WO 2004/016767 A2, Figure 48). Each reaction contained a different biotinylated template (15, 16, 17, 18, 19, or 20) together with five non-biotinylated templates (from 15-20) (WO 2004/016767 A2) and six reagents (21-25) (WO 2004/016767 A2). These reactions were initiated by combining a solution containing 15-20 with a solution containing 21-25. The products that arose from each biotinylated template were captured with streptavidin-coated magnetic beads and identified by denaturing gel electrophoresis. Because the six reagents in each reaction contained oligonucleotides of unique lengths, the formation of any reaction products involving the biotinylated templates and any of the reagents could be detected. In all six cases, the biotinylated template formed only the single product programmed by its DNA sequence (WO 2004/016767 A2, Figure 48) despite the possibility of forming up to five other products in each reaction. Taken together, these findings indicate that reactions of significantly different rates requiring a variety of non-DNA-linked accessory reagents can be directed by DNA-templated synthesis in the same solution, even when both templates and reagents contain several different cross-reactive functional

groups. The ability of DNA templates to direct multiple reactions at concentrations that exclude non-templated reactions from proceeding at appreciable rates mimics, in a single solution, a spatially separated set of reactions.

- Compared to the use of traditional synthetic methods, generating libraries of small molecules by DNA-templated synthesis is limited by several factors including the need to prepare DNA-linked reagents, the restriction of aqueous, DNA-compatible chemistries, and the reliance on characterization methods such as mass spectrometry and electrophoresis that are appropriate for molecular biology-scale (pg to ug) reactions. On the other hand, DNA-templated synthesis (i) allows the direct *in vitro* selection (as opposed to screening) and amplification of synthetic molecules with desired properties, (ii) permits the preparation of synthetic libraries of unprecedented diversity, and (iii) requires only minute quantities of material for selection and identification of active library members. In addition, this example demonstrates that potentially useful modes of reactivity not possible using current synthetic methods can be achieved in a DNA-templated format. For example, six different types of reactions can be performed simultaneously in one solution, provided that required non-DNA-linked accessory reagents are compatible. This reaction mode permits the diversification of synthetic small molecule libraries using different reaction types in a single solution.

Materials and Methods

Synthesis of Templates and Reagents

- Oligonucleotides were synthesized using standard automated solid-phase techniques. Modified phosphoramidites and controlled-pore glass supports were obtained from Glen Research, Sterling, Virginia, USA. Unless otherwise noted, functionalized templates and reagents were synthesized by reacting 5'-H₂N(CH₂)₂ terminated oligonucleotides (for templates) or 3'-OPO₃⁻ CH₂CH(CH₂OH)(CH₂)₂NH₂ terminated oligonucleotides (for reagents) in a 9:1 mixture of aqueous 200 mM pH 7.2 sodium phosphate buffer:DMF containing 2 mg/mL of the appropriate N-hydroxysuccinimide ester (Pierce, Rockford, IL, USA) at 25°C.
- For the aldehyde and nitroalkane-linked oligonucleotides (4b, 4c, 5b, 6a, 17, 24, and 26, Figures 46 and 47, WO 2004/016767 A2) the NHS esters were generated by combining the appropriate carboxylic acid (900 mM in DMF) with equal volumes of dicyclohexylcarbodiimide (900 mM in DMF) and NHS (900 mM in DMF) for 90 minutes. Phosphorane-linked oligonucleotides (2b and 20, Figures 46 and 47, WO 2004/016767 A2) were prepared by a 90 minute reaction of the appropriate amino-terminated oligonucleotide with 0.1 volumes of a 20 mg/mL DMF solution of the NHS ester of iodoacetic acid (SIA, Pierce, Rockford, IL, USA) in pH 7.2 buffer as above, followed by addition of 0.1 volumes of a 20 mg/mL solution of 4-diphenylphosphinobenzonic acid in DMF.
- Thiol-linked template 16 was synthesized by reacting ethylene glycol bis(succinimidylsuccinate) (EGS, Pierce, Rockford, IL, USA) with the appropriate oligonucleotide for 15 minutes, followed by addition of 0.1 volumes of 300 mM 2-

aminoethanethiol. Reagent **5a** was synthesized using 3'-OP0₃-(CH₂)₃SS(CH₂)₃ODMT functionalized controlled-pore glass (CPG) support and reduced prior to use according to the manufacturer's protocol.

The 3'-biotinylated oligonucleotides were prepared using biotin-TEG¹ CPG (Glen Research, Sterling, Virginia, USA). Products arising from biotinylated templates were purified by mixing with 1.05 equivalents of streptavidin-linked magnetic beads (Roche), washing twice with 4 M guanidinium hydrochloride, and eluting with aqueous 10 mM Tris pH 7.6 with 1 mM biotin at 80 °C.

Synthesis of Linkers

10 Linkers between DNA oligonucleotides and the functional groups in **1a-6c** are as follows. **1b** and **1c**: DNA-5'-NH₂; **1a**, **2a-2c**, **3a**, and **3c**: DNA-5'-O(CH₂)₂O(CH₂)₂-NH-; **5a**: DNA-3'-O-(CH₂)₃SH; **4a-4c**, **5b**, **5c**, **6a**, and **6c**: DNA-3'-O-CH₂CH(CH₂OH)(CH₂)₃NH-. Oligonucleotide sequences used to generate all possible products in (WO 2004/016767 A2, Figure 46, lanes 5,9, and 14), with annealing regions underlined: R-TATCTACAGAG-3' [SEQ ID NO: 106]

15 **(1a-1c)**: R-TATCTACAGAGTAGTCT-3' [SEQ ID NO: 107] (**2a-2c**); R-TATCTACAGAGTAGTCTAATGAC-3' [SEQ ID NO: 108] (**3a-3c**); 5'-CAGCCTCTGTAGAT-R [SEQ ID NO: 109] (**4a-4c**); 5'-CTCAGCCTCTGTAGAT-R [SEQ ID NO: 110] (**5a-5c**); 5'-GGCTCAGCCTCTGTAGAT-R [SEQ ID NO: 111] (**6a-6c**). Functionalized templates and reagents were purified by gel filtration (Sephadex G-25) followed by reverse-phase HPLC

20 (0.1 M triethylammonium acetate/acetonitrile gradient). Representative functionalized templates and reagents were further characterized by MALDI mass spectrometry.

Reaction Conditions

All reactions were performed by dissolving reagents and templates in separate vessels in pure water before combining them into a solution of 50 mM aqueous TAPS buffer, pH 8.0, 25 250 mM NaCl at 25 °C for 16 hours with DNA-linked reactants at 60 nM (WO 2004/016767 A2, Figure 47) or at 12.5 nM (WO 2004/016767 A2, Figures 47 and 48). NaBH₃CN, EDC, and sulfo-NHS were present when appropriate as described. Products were analyzed by denaturing polyacrylamide gel electrophoresis using ethidium bromide staining and UV transillumination: Differences in charge states, attached functional groups, and partial secondary structure resulted in modest variations in gel mobility for different functionalized oligonucleotides of the same length (Figures 46-48).

Example 20.

Selection for bifunctional molecules capable of binding to a macromolecular target.

35

This is an example of a selection against 6 protein targets, by affinity selection on immobilized protein (subprocess i). The experiments are described in detail in the patent application (Liu et al., WO 2004/016767 A2, example 11, p. 171-182).

Six proteins, GST, Carbonic anhydrase, Papain, Trypsin, Chymotrypsin, and Streptavidin, were 5 immobilized on NHS-activated Sepharose 4 fast flow beads. For each of the proteins, a known ligand was prepared and linked to a unique DNA sequence. Solutions containing DNA-linked protein ligands and DNA-linked negative controls were used to simulate libraries of bifunctional molecules. The selections were performed by first incubating the DNA-linked ligands with immobilized protein, then beads were washed, and finally the DNA of the DNA-linked ligands that bound to the beads 10 was amplified by PCR, to reveal the efficiency of the model selection experiment. All proteins were enriched more than 50-fold.

Example 21.

15 Iterated selection on immobilized target (subprocess viii).

This is an example of iterated rounds of selection and elution without intervening amplification of the bifunctional molecule (subprocess vii, above). The description of the experiments is taken from (Liu et al., WO 2004/016767 A2, example 11, p. 173). The figures referred to are from the same patent application.

20 Selections can be iterated to multiply the net enrichment of desired molecules. To test this possibility with DNA-linked synthetic molecules, a 1:1,000 mixture of DNA-linked phenyl sulfonamide (3):DNA-linked N-formyl-Met-Leu-Phe (2) (WO 2004/016767 A2) was subjected to a selection for binding carbonic anhydrase. The molecules surviving the first selection were eluted and directly subjected to a second selection using fresh immobilized carbonic anhydrase. PCR 25 amplification and restriction digestion revealed that the first round of selection yielded a 1:3 ratio of (3):(2), representing a 3,30-fold enrichment for the DNA-linked phenyl sulfonamide. The second round of selection further enriched (3) by more than 30-fold, such that the ratio of (3):(2) following two rounds of selection exceeded 10:1 (>10⁴-fold net enrichment). Similarly, three rounds of iterated selection were used to enrich a 1:10⁶ starting ratio of (3):DNA-linked biotin (4) by a factor of 5 x 10⁶ into a solution containing predominantly DNA-linked phenyl sulfonamide (3) (see WO 2004/016767 A2, Figure 81). These findings demonstrate that enormous net enrichments for DNA-linked synthetic molecules can be achieved through iterated selection, and suggest that 30 desired molecules represented as rarely as 1 part in 10⁶ within DNA-templated synthetic libraries may be efficiently isolated in this manner.

35

Example 22.

Stage 2 reactions: Reductive amination, amine acylation, carbon-carbon forming reactions, and organometallic coupling reactions.

This is an example of reactions that can be employed in a stage 2 synthesis. By maintaining a high concentration of molecule fragments (e.g. 10-100 mM), the conditions applied to the templated synthesis hereunder, can be applied to stage 1 synthesis as well, using the same reaction types. The description of the experiments is taken from (Liu et al., WO 2004/016767 A2, example 2, p. 107-112). The figures referred to are from the same patent application.

As described in detail herein, a variety of chemical reactions for example, DNA-templated organometallic couplings and carbon-carbon bond forming reactions can be utilized to construct small molecules.

The ability of DNA-templated synthesis to direct reactions that require a non-DNA-linked activator, catalyst or other reagent in addition to the principal reactants has also been demonstrated herein.

To test the ability of DNA-templated synthesis to mediate such reactions without requiring

structural mimicry of the DNA-templated backbone, DNA-templated reductive aminations between an amine-linked template (1) (WO 2004/016767 A2) and benzaldehyde- or glyoxal-linked reagents

(3) (WO 2004/016767 A2) with millimolar concentrations of sodium cyanoborohydride (NaBH_3CN) at room temperature in aqueous solutions can be performed (see WO 2004/016767 A2, Figure 23A). Significantly, products formed efficiently when the template and reagent sequences were

complementary, while control reactions in which the sequence of the reagent did not complement that of the template, or in which NaBH_3CN was omitted, yielded no significant product (see WO

2004/016767 A2, Figures 23A-23D and 24). Although DNA-templated reductive aminations to generate products closely mimicking the structure of double-stranded DNA have been previously reported (see, for example, Li et al. (2002) J. Am. Chem. Soc. 124: 746 and Gat et al. (1998) Biopolymers 48:19), these results demonstrate that reductive amination to generate structures

unrelated to the phosphoribose backbone can take place efficiently and sequence-specifically.

Referring to (WO 2004/016767 A2, Figures 25A-25B, DNA-templated amide bond formations between amine-linked templates 4 and 5 and carboxylate-linked reagents 6-9 mediated by 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide (EDC) and N-hydroxysulfosuccinimide (sulfo-NHS) generated amide products in good yields at pH 6.0, 25°C. Product formation was (i) sequence-specific, (ii) dependent on the presence of EDC, and (iii) insensitive to the steric encumbrance of the amine or carboxylate. Efficient DNA-templated amide formation was also mediated by the water-stable activator 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride (DMT-MM). Instead of EDC and sulfo-NHS (WO 2004/016767 A2, Figures 24 and 25A-25B). The efficiency and generality of DNA-templated amide bond formation under these conditions, together

with the large number of commercially available chiral amines and carboxylic acids, make this reaction an attractive candidate in future DNA-templated syntheses of structurally diverse small molecule libraries.

Carbon-carbon bond forming reactions are also important in both chemical and biological syntheses and thus several such reactions can be utilized in a nucleic acid-templated format. Both the reaction of nitroalkane-linked reagent (10) (WO 2004/016767 A2) with aldehyde-linked template (11) (WO 2004/016767 A2) (nitro-aldol or Henry reaction) and the conjugate addition of

5 10 to maleimide-linked template (12) (WO 2004/016767 A2) (nitro-Michael addition) proceeded efficiently and with high sequence specificity at pH 7.5-8.5, 25°C (WO 2004/016767 A2, Figures 23A and 24). In addition, the sequence-specific DNA-templated Wittig reaction between stabilized phosphorus ylide reagent 13 (WO 2004/016767 A2) and aldehyde-linked templates 14 or 11 (WO

2004/016767 A2) provided the corresponding olefin products in excellent yields at pH 6.0-8.0, 25°C (WO 2004/016767 A2, Figures 23B and 24). Similarly, the DNA templated 1,3-dipolar

cydoaddition between nitroene-linked reagents 15 and 16 (WO 2004/016767 A2) and olefin-linked templates 12, 17 or 18 also afforded products sequence specifically at pH 7.5, 25°C (WO 2004/016767 A2, Figures 23B, 23C and 24).

In addition to the reactions described above, organometallic coupling reactions can also be utilized

15 in the present invention. For example, DNA-templated Heck reactions were performed in the presence of water-soluble Pd precatalysts. In the presence of 170 mM Na_2PdCl_4 , aryl iodide-linked

reagent 19 (WO 2004/016767 A2) and a variety of olefin-linked templates including maleimide 12, acrylamide 17, vinyl sulfone 18 or dinnamamide 20 (WO 2004/016767 A2) yielded Heck coupling

products in modest yields at pH 5.0, 25°C (WO 2004/016767 A2, Figures 23D and 24). For

20 couplings with olefins 17, 18 and 20, adding two equivalents of $\text{P}(\text{p-SO}_3\text{C}_6\text{H}_4)_3$ per equivalent of Pd prior to template and reagent addition typically increased overall yields by 2-fold. Control reactions containing sequence mismatches or lacking Pd precatalyst yielded no product.

In order to evaluate the ability of the DNA-templated reactions to take place efficiently when

25 reactants are separated by distances relevant to library encoding, the yields of reductive

amination, amide formation, nitro-aldol addition, nitro-Michael addition, Wittig olefination, dipolar

cydoaddition, and Heck coupling reactions were compared when either zero ($n \sim 0$) or ten ($n =$

10) bases separated the annealed reactive groups. Among the reactions described here, amide

bond formation, nitro-aldol

addition, Wittig olefination, Heck coupling, conjugate addition of thiols to maleimides and $\text{S}_{\text{N}}2$

30 reaction between thiols and alpha-iodo amides demonstrate comparable product formation when

reactive groups are separated by zero or ten bases (WO 2004/016767 A2, Figure 26B). Figure

26B shows the results of denaturing polyacrylamide gel electrophoresis of a DNA-templated Wittig

olefination between complementary 11 and 13 with either zero bases (lanes 1-3) or ten bases

(lanes 4-6) separating the annealed reactants. Product yields after 13 hours at both distances

35 were nearly quantitative.

Control reactions containing sequence mismatches yielded no detectable product. These findings indicate that these reactions can be encoded during synthesis by nucleotides that are distal from the reactive end of the template without significantly impairing product formation.

In addition to the DNA-templated S_N2 reaction, conjugate addition, vinyl sulfone addition, amide bond formation, reductive amination, nitro-alcohol (Henry reaction), nitro Michael, Wittig olefination, 1,3-dipolar cycloaddition and Heck coupling reactions described above, a variety of additional reagents can also be utilized in the method of the present invention. For example, as depicted in (WO 2004/016767 A2, Figure 27), powerful aqueous DNA-templated synthetic reactions including, but not limited to, the Lewis acid-catalysed aldol addition, Mannich reaction, Robinson annulation reactions, additions of allyl indium, zinc and tin to ketones and aldehydes, Pd-assisted allylic substitution, Diels-Alder cycloadditions, and hetero-Diels-Alder reactions can be utilized efficiently in aqueous solvent and are important complexity-building reactions.

10 Taken together, these results expand considerably the reaction scope of DNA-templated synthesis. A wide variety of reactions can proceed efficiently and selectively when the corresponding reactants are programmed with complementary sequences. By augmenting the repertoire of known DNA-templated reactions to include carbon-carbon bond forming and organometallic reactions (nitro-alcohol additions, nitro-Michael additions, Wittig olefinations, dipolar cycloadditions, and Heck couplings, in addition to previously reported amide bond formation (see, Schmidt *et al.* (1997) *Nucleic Acids Res.* 25:4792; Bruck *et al.* (1996) *Chem. Biol.* 3: 49), imine formation (Czapinski *et al.* (2001) *J. Am. Chem. Soc.* 123: 8618), reductive amination (Liu *et al.* (2002) *J. Am. Chem. Soc.* 124: 746; Gat *et al.* (1998) *BIOPOLYMERS* 48:19), S_N2 reactions (Gartner *et al.* (2001) *J. Am. Chem. Soc.* 123: 6961; Xu *et al.* (2001) *Nat. Biotechnol.* 19: 148; Herlein *et al.* (1995) *J. Am. Chem. Soc.* 117: 10151), conjugate addition of thiols (Gartner *et al.* (2001) *J. Am. Chem. Soc.* 123: 6961), and phosphoester or phosphoramidate formation (Orgel *et al.* (1995) *Acc. Chem. Res.* 28: 109; Luther *et al.* (1998) *Nature* 396: 245), these results may permit the sequence-specific translation of libraries of DNA into libraries of structurally and functionally diverse synthetic products.

25 Because minute quantities of templates encoding desired molecules can be amplified by PCR, the yields of DNA-templated reactions arguably are less critical than the yields of traditional synthetic transformations. Nevertheless, many of the reactions discussed in this example proceed efficiently.

Materials and Methods

30 Functionalized templates and reagents were typically prepared by reacting 5' -NH₂ terminated oligonucleotides (for template 1), 5' -NH₂ - (CH₂)₂ terminated oligonucleotides (for all other templates) or 3' -OP(=O)(CH₂)₂CH₂OH(CH₂)₂NH₂ terminated nucleotides (for all reagents) with the appropriate NHS esters (0.1 volumes of a 20 mg/mL solution in DMF) in 0.2 M sodium phosphate buffer, pH 7.2, 25°C, for 1 hour to provide the template and reagent structures shown in (WO 2004/016767 A2, Figures 23A-23D and 25A-25B). For amino acid linked reagents 6-9, 3' - OP(=O)(CH₂)₂CH₂OH(CH₂)₂NH₂ terminated oligonucleotides in 0.2 M sodium phosphate -buffer, pH 7.2 were reacted with 0.1 volumes of a 100 mM bis[2-(succinimidyl)oxycarbonyloxy]ethyl/sulfone

(BSOCIES, Pierce, Rockford, IL, USA) solution in DMF for 10 minutes at 25°C, followed by 0.3 volumes of a 300 mM amino acid in 300 mM sodium hydroxide (NaOH) for 30 minutes at 25°C.

Functionalized templates and reagents were purified by gel filtration using Sephadex G-25 followed by reverse-phase HPLC (0.1 triethylammonium acetate-acetonitrile gradient) and characterized by MALDI mass spectrometry. For the DNA templated reactions described in (WO 2004/016767 A2, Figures 23A-23D) reactions were conducted at 25 °C with one equivalent each of template and reagent at 60 nM final concentration unless otherwise specified. Conditions: (a) 3mMNaBH₃CN, 0.1 M V-[2-morpholinoethane] sulfonic acid (MES) buffer pH 6.0, 0.5 M NaCl, 1.5 hours; b) 0.1 M *N*-tris[hydroxymethyl] methyl-3-aminopropanesulfonic acid (TAPS) buffer pH 8.5, 300 mM NaCl, 12 hours; c) 0.1 M pH 8.0 TAPS buffer, 1 M NaCl, 5°C, 1.5 hours; d) 50 mM MOPS buffer pH 7.5, 2.8 M NaCl, 22 hours; e) 120 nM 19, 1.4 mM Na₂PdCl₄, 0.5 M NaOAc buffer pH 5.0, 18 hours; (f) Premix Na₂PdCl₄ with two equivalents of P(p-SO₃CH₃)₃ in water for 15 minutes, then add to reactants in 0.5 M NaOAc buffer pH 5.0, 75 mM NaCl, 2 hours (final [Pd] = 0.3 mM, [19] = 120 nM). The olefin geometry of products from 13 and the regiochemistries of cycloaddition products from 14 and 16 are presumed but not verified (WO 2004/016767 A2, Figures 23A-23D). Products were characterized by denaturing polyacrylamide gel electrophoresis and MALDI mass spectrometry. For all reactions under the specified conditions, product yields of reactions with matched template and reagent sequences were greater than 20-fold higher than that of control reactions with scrambled reagent sequences.

The conditions for the reactions described in (WO 2004/016767 A2, Figures 25A-25B) were: 60 nM template, 120 nM reagent, 50 mM DMT-MM in 0.1 M MOPS buffer pH 7.0, 1 M NaCl, for 16 hours at 25°C; or 60 nM template, 120 nM reagent, 20 mM EDC, 15 mM sulfo-NHS, 0.1 M MES buffer pH 6.0, 1 M NaCl, for 16 hours at 25°C. In each row of the table in (WO 2004/016767 A2, Figures 25A-25B), yields of DMT-MM-mediated reactions between reagents and templates complementary in sequence were followed by yields of EDC and sulfo-NHS-mediated reactions. In all cases, control reactions with mismatched reagent sequences yielded little or no detectable product and products were characterized by denaturing polyacrylamide gel electrophoresis and MALDI mass spectrometry.

30 (WO 2004/016767 A2, Figure 24) depicts the analysis by denaturing polyacrylamide gel electrophoresis of representative DNA-templated reactions listed in (WO 2004/016767 A2, Figures 23A-23D and 25A-25B). The structures of reagents and templates correspond to the numbering in Figures 23A-23D and 25A-25B. Lanes 1, 3, 5, 7, 9, 11: reaction of matched (complementary or "M") reagents and templates under conditions listed in Figures 23A-23D and 25A-25B (the reaction between 4 and 6 was mediated by DMT-MM). Lanes 2, 4, 6, 8, 10, 12: reaction of mismatched (non-complementary or "X") reagents and templates under conditions identical to those in lanes 1, 3, 5, 7, 9 and 11, respectively.

The sequences of oligonucleotide templates and reagents are as follows (5' to 3' direction, *n* refers to the number of bases between reactive groups when template and reagent are annealed as shown in (WO 2004/016767 A2, Figure 26A).

- 5 1: TGGTAGCAATTCGACTCGGG [SEQ ID NO: 45];
 2 and 3 matched: GAGTCGAATTCGTACC [SEQ ID NO: 46];
 2 and 3 mismatched: GGGCTCAGCTCCCA [SEQ ID NO: 47];
 4and5: GGTACGAATTCGACTCGGAAATACACCTT [SEQ ID NO: 48];
 6-9 matched (*n* = 10): TCCCGAGTCG [SEQ ID NO: 49];
 10 6 matched (*n* = 0): AATTCTGTACC [SEQ ID NO: 50];
 6-9 mismatched: TCACCTAGCA [SEQ ID NO: 51];
 11,12,14,17,18, 20: GGTACGAATTCGACTCGGGA [SEQ ID NO: 52];
 10,13,16,19 matched: TCCCGAGTCGAATTCGTACC [SEQ ID NO: 53];
 10,13,16,19 mismatched: GGCTCAGCTCCCAATAT [SEQ ID NO: 54];
 15 15 matched: AATTCTGTACC [SEQ ID NO: 55];
 15 mismatched: TCGATTTCCA [SEQ ID NO: 56];
 template for *n* = 10 vs. *n* = 0 comparison: TAGCGATTCGAGTAATTCGACTCGGGA [SEQ ID NO: 57].

Reaction yields were quantitated by denaturing PAGE followed by ethidium

- 20 bromide staining, UV visualization, and charge-coupled device (CCD)-based densitometry of product and template starting material bands. Yield calculations assumed that templates and 25 products stained with equal intensity per base; for those cases in which products were partially double-stranded during quantitation, changes in staining intensity may have resulted in higher apparent yields.

25

Example 23.

Different stage 1 and stage 2 synthesis schemes employed in a given series of experiments.

- 30 Because of the modular nature of the stage 1, stage2 and selection/screening protocols, it is perfectly possible to generate a first generation library using e.g. subprocess (1, i.e., no templated synthesis involved), then select (e.g. using subprocess 1), and then perform a second round of library generation, and this time use the recovered templates as templates, and therefore, perform a stage 2 synthesis to make the enriched second generation library. Obviously, it is important to 35 keep the same code for the same molecule fragments,
 It may also be advantageous to select against immobilized target in the first round, and then in the second round perform in solution selection experiments for example, or some other selection experiment that share few of the same features as the first selection assay.

Example 24.

Carrier preparation by several different routes.

- 5 Because of the modular nature of the stage 1 synthesis procedures, the carrier that are employed in a stage 2 synthesis can be prepared by different synthetic routes. As an example, in order to make e.g. 2.000 Identifiers, with the ability to make 1.000.000 different template-encoded molecules, one could synthesize 625 carriers by two step Lerner-like stage 1 synthesis (subprocess 1), using acylation reactions to link the molecule 10 fragments; synthesize 1000 carriers using the DNA-routing approach by Harbury (subprocess 10), for example employing reductive amination and nucleophilic aromatic substitution reactions; synthesize 375 compounds by combinatorial chemistry and attach these to identifiers. Then use this pool of 2000 carriers in a stage 2 synthesis to generate 1.000.000 bifunctional molecules.

15

Example 25.

Stage 1 synthesis employing the Harbury and Halpin method (subprocess 10).

- 20 Subprocess 10 stage 1 synthesis involves a DNA sorting step, in which the identifiers to be linked to the molecule fragments are sorted according to their DNA sequences. Once the DNA has been sorted, the molecule fragments can be linked under conditions identical to the conditions described in the present invention, in particular, as described in all of the above examples. Thus, the preferred reactions, reductive amination, Wittig reaction, 25 acylation, alkylhalide alkylation, nucleophilic aromatic substitution, Heck coupling, cycloaddition reactions, sulfonylation, isocyanide addition, Michael addition and others, may be executed in exactly the same way as described here.

Applications of the present invention.

- 30 The methods of the present invention provide for the identification of organic and inorganic molecules that are catalysts useful for the synthesis of complex molecules from simple substrates, inorganic compounds with useful properties as materials, may be used in the degradation of plastics, animal feed processing, etc. Also, the methods can be applied to identification of compounds with high affinity or selectivity for targets and surfaces, 35 including protein targets, DNA, and other macromolecular structures, metal surfaces, plastics, etc. Such compounds may be useful as additives to paint, cement, textiles, and other substances where improved rigidity, strength, flexibility or stability is desired. New materials may be identified in this way, including superconductors and nanosensors.

Compounds that bind with high affinity and/or selectivity to protein, RNA, DNA, polysaccharides, or other molecules of an organism, may be used in diagnostics or as therapeutics.

5

REFERENCES

- Balndur et al., US 5,891,737
- Balndur et al., US 5,646,285
- Boger et al., US 6,194,612 B1
- Cook et al., US 6,191,273
- Dervan et al., US 6,090,947;
- Dower et al., US 1991000762522
- Dower et al., EP 0604552 B1
- Franch et al., WO 2004/083427
- 15 Freskgård et al., WO 2004/039825 A2
- Graybill et al., US 6,127,191
- Gustafson et al., US 6,140,361
- Harbury and Halpin, WO 00/23458
- Lebl et al., US 6,090,912
- 20 Lebl et al., US 5,840,485
- Lerner, R. et al., EP 0643778 B1
- Liu et al., WO02/074929A2
- Liu et al., WO 2004/016767 A2
- Morgan et al., WO 2005/058479
- 25 Pedersen et al., WO02/103008A2
- Pedersen et al., WO03/078625 A2
- Still et al. (1998) US 5,721,099;
- Abelson (1996) *Methods in Enzymology*, 267; 211-221;
- Alvarez et al. (1999), *Journal of Organic Chemistry*, 64; 6319-28;
- 30 Ashley, G.W., Kushlan, D.M. (1991), *Biochemistry*, 30; 2927-2933
- Bruick et al. (1996), *Chem. Biol.* 3; 49
- Calderone et al. (2002), *Angew. Chem. Int. Ed. Engl.*, 41; 4104;
- Czapinski et al. (2001), *J. Am. Chem. Soc.*, 123; 8618
- Dapremont et al. (1995) *Physiol. Chem. Phys. & Med. NMR*, 27; 339-343,
- 35 Dollinnaya, N.G. et al. (1993), *Nucl. Acids. Res.*, 21; 5403-5407
- Dollinnaya, N.G., et al. (1994), *Nucleosides Nucleotides* 13; 2169-2183
- Doyon et al. (2003), *J. Am. Chem. Soc.*, 125; 12372-12373
- Feuston et al. (2002), *J. Med. Chem.*, 45; 5640-8

- Fruchart et al. (1999), *Tetrahedron Lett.*, 40; 6225
- Gao, H., et al. (1994), *Bioconjugate Chem.*, 5; 445-453
- Gartner et al. (2001), *J. Am. Chem. Soc.* 123; 6961
- Gartner et al. (2002), *Angew. Chem. Int. Ed.* 41, 1796
- 5 Gat et al. (1998), *Biopolymers*, 48; 19
- Goodwin, J.T., and Lynn, D.G. (1992), *J. Am. Chem. Soc.*, 114; 9197-9198)
- Hertlein et al. (1995), *J. Am. Chem. Soc.*, 117; 10151
- Hughes (1996), *Tetrahedron Lett.*, 37; 7595
- Kool, E.T. (1991), *J. Am. Chem. Soc.*, 113; 6265-6266
- 10 Kramer et al. (1999), *Current Protocols in Mol. Biol.*, 3; 15.1
- Jost et al. (1989), *Nucleic Acids Res.*, 17; 2143
- Liel et al. (2002), *J. Am. Chem. Soc.*, 124; 746
- Luebke, et al. (1991), *J. Am. Chem. Soc.*, 113; 7447-7448
- Luebke et al. (1992), *Nucl. Acids Res.*, 20; 3005-3009
- 15 Luther et al. (1998), *Nature*, 396; 245
- Lynn et al. (1998), *J. Am. Chem. Soc.*, 120; 1627-1628
- Lynn et al. (2000), *J. Am. Chem. Soc.*, 122; 6601-6609
- Melnikov et al. (1999), *Langmuir*, 15; 1923-1928
- Mohr et al. (1996) *Organometallics*, 15; 4317- 4325
- 20 Orgel et al. (1995), *Acc. Chem. Res.*, 28; 109
- Plrung et al. (1998), *Journal of Organic Chemistry*, 63; 241-46
- Rohatgi, R., et al. (1996), *J. Am. Chem. Soc.*, 118; 3332-3339
- Schmidt et al (1997), *Nucleic Acids Res.*, 25; 4792
- Shabarova, Z.A. (1988), *Biochemie*, 70; 1323-1334;
- 25 Sokolova, N.I. et al. (1988), *FEBS Letters*, 232; 153-155;
- Valisman et al., (2001) *J. Biol. Chem.*, 276; 30615-30622
- Vischer, J., Shwartz, A.W. (1988), *Journal of Molecular Evolution*, 28; 3-6
- Walder et al. (1979), *Proc. Natl. Acad. Sci.*, 76; 51-55
- Wang, E., Yanagawa, H. (1986), *Biochemistry*, 25;7423-7430
- 30 Washington et al., (2001) *J. Biol. Chem.*, 276; 2263-2266
- Yanzheng, X., and Kool, E.T. (1997), *Tetrahedron Letters*, 38; 5595-5598
- Xu et al. (2001), *Nat. Biotechnol.*, 19; 148
- Zhang et al. (1999), *J. Am. Chem. Soc.* 121; 3311
- Zhao et al. (1998), *Journal of Organic Chemistry*, 63; 7568-7572
- 35 Zarling et al., (1980), *J. Immunol.*, 124; 913

CLAIMS

1. A method for synthesizing an encoded molecule comprising the steps of:

- h) Adding a linker molecule L to one or more reaction wells;
- i) Adding a molecule fragment to each of said reaction wells;
- j) Adding an oligonucleotide identifier to each of said reaction wells;
- k) Subjecting said wells to conditions sufficient to allow said molecule fragments and said oligonucleotide identifiers to become attached to said linker molecule, or conditions sufficient for said molecule fragments to bind to other molecule fragments and sufficient for said oligonucleotide identifiers to bind to other oligonucleotide identifiers;
- l) Combining the contents of said one or more reaction wells;
- m) Optionally, distributing the combined product to one or more new reaction wells;
- n) Optionally, repeating steps b) to e) one or more times;
- o) Contacting the resulting bifunctional molecule(s) of step e) or g) with one or more templates each capable of hybridizing to at least one of the oligonucleotide identifiers added in step c);

15 wherein

the linker molecule L contains at least one reactive group capable of reacting with a reactive group in the molecule fragment and at least one reactive group capable of reacting with a reactive group in the oligonucleotide;

20 the molecule fragments each contain at least one reactive group capable of reacting with a reactive group in the linker molecule L or a reactive group in another molecule fragment, and the reactive groups of each molecule fragment may be the same or different;

25 the oligonucleotide identifiers each contain at least one reactive group capable of reacting with a reactive group in the linker L or a reactive group in another oligonucleotide identifier, and the reactive groups of each oligonucleotide identifier may be the same or different;

30 the oligonucleotide identifier added to each well in step c) identifies the molecule fragment added to the same well in step b);

the steps a) to d) may be performed in any order;

the steps b) to d) in step f) may also be performed in any order;

the number of wells in steps a) and f) may be the same or different;

the oligonucleotide template optionally is associated with a reactive group.

35 2. A method according to claim 1, wherein reactive groups of a molecule fragment of each of two or more bi-functional molecules hybridized to the same template are reacted, or wherein the reactive group of a molecule fragment of one bi-functional molecule is reacted with a reactive group associated with the template to which it is hybridized.

3. The method according to claim 1 or 2, wherein the number of wells in step a) is m and the number of wells in step f) is n, and wherein for each repetition of steps b) to d) in step f) n may be the same or different; and the structure of the encoded molecule is

$$5 \quad O_{p,q} - (((O_{2,q} - (O_{1,q} - Z))) - Y - (((X - R_{1,q}) - R_{2,q}) - \dots) - R_{p,q})$$

wherein

X, Y and Z are components of the linker molecule, L, X being adapted for reaction with a molecule fragment, Z being adapted for reaction with an oligonucleotide and Y being a flexible linker connecting X and Z;

10 $O_{p,q}$ is the oligonucleotide identifier added in repetition number (p-1) of steps b) to f) in well number q;

$R_{p,q}$ is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

p is an integer of at least 2;

15 m and n are integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and

for $O_{p,q}$ and $R_{1,q}$, q is in the range 1 to m, for $O_{p,q}$ and $R_{p,q}$ where p is greater than 1, q is in the range 1 to n.

20 4. The method according to any of the preceding claims, wherein the number of wells in step a) is m and the number of wells in step f) is n, and wherein for each repetition of steps b) to d) in step f) n may be the same or different; and the structure of the encoded molecule is

$$O_{p,q} - (((O_{2,q} - O_{1,q}) - ((L - R_{1,q}) - R_{2,q}) - \dots) - R_{p,q})$$

25 wherein

$O_{p,q}$ is the oligonucleotide identifier added in repetition number (p-1) of steps b) to f) in well number q;

$R_{p,q}$ is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

30 p is an integer of at least 1;

m and n are integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and

for $O_{1,q}$ and $R_{1,q}$, q is in the range 1 to m, for $O_{p,q}$ and $R_{p,q}$ where p is greater than 1, q is in the range 1 to n.

35

5. A method for synthesizing an encoded molecule comprising the steps of:

- j) Dispensing aliquots of a nascent linker molecule L into each of m reaction wells;

- k) Dispensing into each of said m reaction wells a corresponding aliquot of an mth molecule fragment, R_{1,m} and a corresponding aliquot of an mth oligonucleotide, O_{1,m} Identifier;
- l) Combining all of the nascent bi-functional molecules from all m reaction wells to produce an admixture of nascent bi-functional molecules;
- m) Optionally, Dispensing said admixture of nascent bi-functional molecules into n reaction wells
- n) Optionally, dispensing into each of the n reaction wells of step d) a corresponding aliquot of an mth molecule fragment, R_{p,q}, and a corresponding aliquot of an mth oligonucleotide Identifier, O_{p,q};
- o) Optionally, combining all of the nascent bi-functional molecules from all m reaction wells in step e) to produce an admixture of nascent bi-functional molecules;
- p) Optionally repeating steps d) to f) one or more times;
- q) contacting a resulting bi-functional molecule of step f) or g) with one or more templates, said one or more templates optionally being associated with a reactive group, under conditions to allow for hybridization of each of the templates to one or more of said nascent bi-functional molecule generated in step f) or g);
- r) Optionally, reacting reactive groups of a molecule fragment of two or more nascent bi-functional molecules hybridized to the same template, or reacting the reactive group of a molecule fragment of one nascent bi-functional molecule with the reactive group associated with the template to which it is hybridized;

Wherein:

the linker molecule L contains at least one reactive group capable of reacting with a reactive group in the molecule fragment and at least one reactive group capable of reacting with a reactive group in the oligonucleotide;

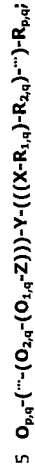
the molecule fragments each contain at least one reactive group capable of reacting with a reactive group in the linker molecule L or a reactive group in another molecule fragment, and the reactive groups of each molecule fragment may be the same or different;

the oligonucleotide identifiers each contain at least one reactive group capable of reacting with a reactive group in the linker L or a reactive group in another oligonucleotide Identifier, and the reactive groups of each oligonucleotide Identifier may be the same or different;

the oligonucleotide Identifier added to each well in step b) and e) identifies the molecule fragment added to the same well in the respective step;

the steps a) and b) as well as the steps d) and e) may be performed in any order; the steps d) and e) in step f) may also be performed in any order;

6. The method according to claim 5, wherein the number of wells in step a) is m and the number of wells in step f) is n, and wherein for each repetition of steps b) to d) in step f) n may be the same or different; and the structure of the encoded molecule is



wherein

X, Y and Z are components of the linker molecule, L, X being adapted for reaction with a molecule fragment, Z being adapted for reaction with an oligonucleotide and Y being a flexible linker connecting X and Z;

10 O_{p,q} is the oligonucleotide Identifier added in repetition number (p-1) of steps b) to f) in well number q;

R_{p,q} is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

p is an Integer of at least 2;

15 m and n are Integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and for O_{1,q} and R_{1,q}, q is in the range 1 to m, for O_{p,q} and R_{p,q} where p is greater than 1, q is in the range 1 to n.

20 7. The method according to claim 5 or 6, wherein for each repetition of steps d) to f) in step g) n may be the same or different; and the structure of the encoded molecule is



wherein

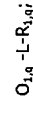
25 O_{p,q} is the oligonucleotide Identifier added in repetition number (p-1) of steps b) to f) in well number q;

R_{p,q} is the molecule fragment added in repetition number (p-1) of steps b) to f) in well number q;

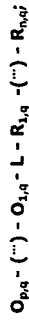
p is an Integer of at least 1;

30 m and n are Integers of at least 5, such as at least 10, preferably at least 15, more preferably at least 20, and most preferably at least 50; and for O_{1,q} and R_{1,q}, q is in the range 1 to m, for O_{p,q} and R_{p,q} where p is greater than 1, q is in the range 1 to n.

8. The method according to any of claims 5 to 7, wherein the structure of the nascent bi-functional molecules resulting from step b) is



and the structure of the nascent bi-functional molecule obtained after repeating the process steps defined in step g) p-1 times is



and wherein p is greater than or equal to 1.

5

9. The method according to any of claims 5 to 8, wherein the number number of reaction wells in step a) and/or the number of reaction wells in step f) is at least 2, such as at least 5, at least 10, at least 25, at least 50, at least 100, at least 200, at least 500, at least 1000 at least, at least 10,000, at least 100,000 at least 1,000,000, at least 10^7 , or at least 10^8 .

10

10. The method according to any of the preceding claims, wherein the process steps defined in step g) are repeated at least once, such as at least twice, such as at least three times, such as at least four times.

15

11. The method according to any of claims 3 to 8, wherein the oligonucleotide identifier, $O_{p,q}$, added in reaction well number q in repetition number (p-1) of the steps specified in step g) uniquely identifies the molecule fragment, $R_{p,q}$, added in reaction well number q in repetition number (p-1) of the steps specified in g).

20

12. The method according to any of claims 3 to 8, wherein identical oligonucleotide identifiers, $O_{p,q}$, are added in two or more reaction wells in repetition number (p-1) of the steps specified in g).

25 13. The method according to any of claims 3 to 8, wherein identical molecule fragments, $R_{p,q}$, are added in two or more wells in repetition number (p-1) of the steps specified in g).

14. The method according to any of claims 3 to 8 and 10 to 11, wherein two or more oligonucleotide identifiers are added to one or more reaction wells in a repetition of the steps specified in g).

30

15. The method according to any of claims 3 to 8 and 10 to 12, wherein two or more molecule fragments are added to one reaction wells in each repetition of the steps specified in g).

35

16. The method according to any of the preceding claims, wherein each oligonucleotide identifier comprises a sequence of from 2 to 100, such as from 2 to 90, from 2 to 80, from 2 to 70, from 2 to 60, from 2 to 50, from 2 to 40 from 2 to 30, from 2 to 25, from 2 to 20, from 2 to 15, from 2 to 10 or from 2 to 5 nucleotides, such as at least 2, 3, 4, 5,

6,7,8,9,10, 11, 12, 13, 14, 15, 20, 25, 30, 35 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100 nucleotides.

17. The method according to any of the preceding claims wherein the number of identifiers in each resulting bi-functional molecule is 2.

5

18. The method according to any of the preceding claims wherein each template is capable of hybridising with 2 carriers.

19. The method according to any of the preceding claims wherein the number of nucleotides in each oligonucleotide identifier is 10-15.

10

20. The method according to any of the preceding claims wherein said one or more templates comprise a sequence selected from the group consisting of: nucleotides, unnatural nucleotides, PNA, morpholinos, LNA, RNA, DNA, and other nucleotide analogs capable of base pairing with a natural oligonucleotide or unnatural oligonucleotide.

15

21. The method according to any of the preceding claims, wherein said one or more templates have a length of at least 80 nucleotides.

22. The method according to any of the preceding claims, wherein each oligonucleotide identifier and/or a sequence of two or more of the oligonucleotide identifiers comprises a sequence of oligonucleotides, the complementary sequence of which is at least 20% identical to the encoding part of said template.

25

23. The method according to any of the preceding claims, wherein said linker, L, is selected from the group consisting of: Polyethylen glycol, polypeptide, polysaccharide, oligonucleotide, C_6 , C_8 , and C_{12} .

24. The method according to any of the preceding claims, wherein said linker, L, is a cleavable linker.

30

25. The method according to any of the preceding steps, wherein a further step is added comprising contacting the bi-functional molecule resulting from step h) with one or more nascent bi-functional molecules and one or more templates each capable of recognizing at least two of the oligonucleotide identifiers present in the bi-functional molecule.

35

26. The method according to any of the preceding claims, wherein two or more oligonucleotide identifiers are covalently linked together, optionally in the presence of a ligase or isomerase enzyme.

40

27. The method according to any of the preceding claims wherein two or more oligonucleotide identifiers are linked by templated extension by enzymes.

28. The method according to any of the preceding claims wherein two or more oligonucleotide identifiers are ligated by chemical ligation or by enzymatic ligation in one process step and chemical ligation in a preceding or subsequent process step.

5 29. The method according to any of the preceding claims, wherein the linker of at least one bi-functional molecule is cleaved simultaneously with or subsequently to hybridisation of the oligonucleotide identifier of said nascent bi-functional molecule to the template.

30. The method according to any of the preceding claims, wherein said reactive groups are
10 reacted in a reaction selected from the group consisting of: acylation, reductive amination, alkylhalide alkylation, Wittig reaction, sulphonylation, isocyanate addition, carbamate coupling, nucleophilic aromatic substitution, thiourea bond formation, carbamate formation, Heck coupling, HWE reaction, 1,3-dipolar cycloaddition, Michael addition, nitro
aldol condensation.

15

31. The method according to any of the preceding claims, wherein the resulting encoded molecule has a linear structure and is selected from the group consisting of: dimers, trimers, tetramers, pentamers, multimers, and polymers.

20 32. The method according to any of claims 1 to 5 and 7 to 30 wherein at least one molecule fragment having more than one reactive groups has been used in the preparation of the resulting encoded molecule.

33. A method for identifying a molecule with desired characteristics, said method
25 comprising synthesizing a library of encoded molecules by a method as described in any of claims 1-32.

34. A method according to claim 33 further comprising a step of subjecting the library to a partitioning or enrichment procedure, to identify encoded molecules with desired
30 characteristics.

35. A method according to claim 33 or 35, said method further comprising screening a plurality of encoded molecules in a process to identify and optionally increase the relative amount of an encoded molecule having one or more desired characteristics.

35

36. A method according to any of claims 33 to 35, said method comprising identifying the encoded molecule by determining the oligonucleotide sequence(s) of the attached identifiers.

40 37. A library of encoded molecules obtained by a method according to any of claims.

38. A library according to claim 37, wherein the number of different compounds is at least
100.

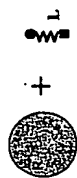
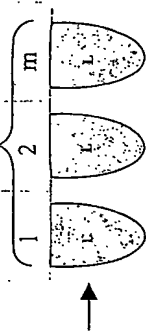
5

10

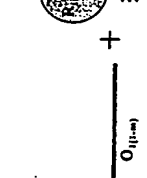
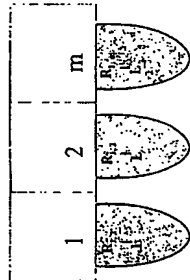
15

20

25



=



=

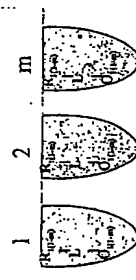
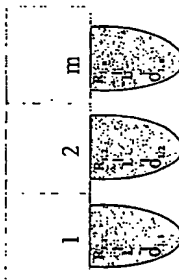
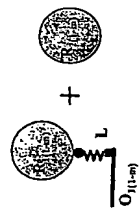
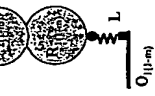


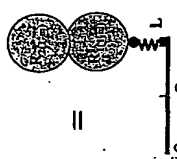
Fig. 1



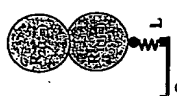
+



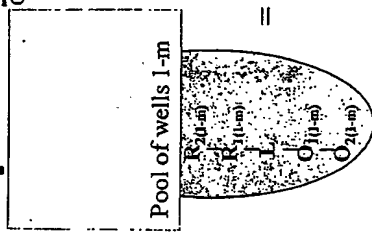
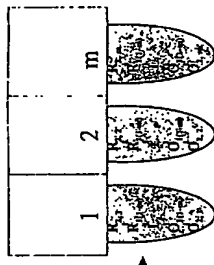
=



+



=



$$\begin{aligned}
 & \left(\begin{matrix} R_{1,1} \\ R_{1,1} \\ L \\ O_{1,1} \end{matrix} \right) + \left(\begin{matrix} R_{2,1} \\ R_{1,1} \\ L \\ O_{1,1} \end{matrix} \right) + \dots + \left(\begin{matrix} R_{1,m} \\ R_{1,1} \\ L \\ O_{1,m} \end{matrix} \right) \\
 & + \left(\begin{matrix} R_{2,2} \\ R_{1,1} \\ L \\ O_{1,1} \end{matrix} \right) + \left(\begin{matrix} R_{2,2} \\ R_{1,2} \\ L \\ O_{1,2} \end{matrix} \right) + \dots + \left(\begin{matrix} R_{2,m} \\ R_{1,2} \\ L \\ O_{1,2} \end{matrix} \right) \\
 & + \left(\begin{matrix} R_{2,1} \\ R_{1,m} \\ L \\ O_{1,m} \end{matrix} \right) + \left(\begin{matrix} R_{2,2} \\ R_{1,m} \\ L \\ O_{1,m} \end{matrix} \right) + \dots + \left(\begin{matrix} R_{2,m} \\ R_{1,m} \\ L \\ O_{1,m} \end{matrix} \right)
 \end{aligned}$$

Fig. 1 cont.

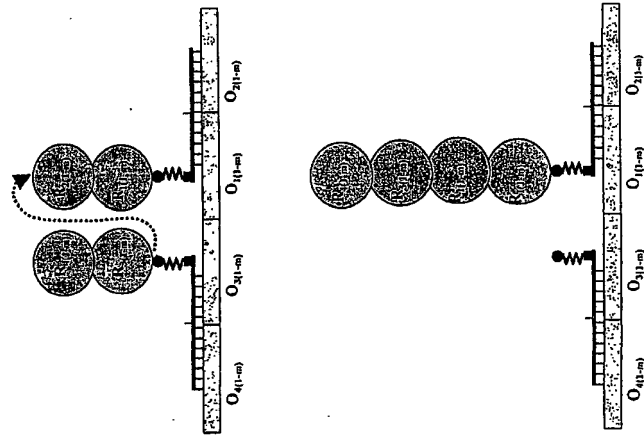
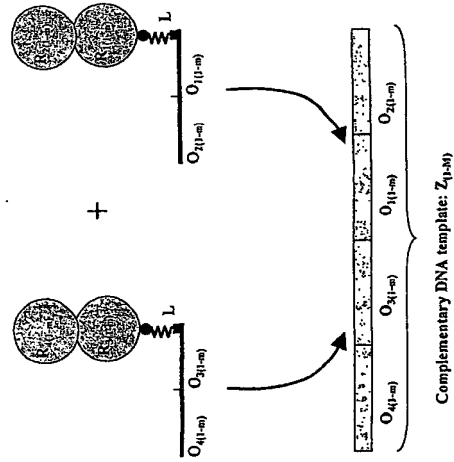
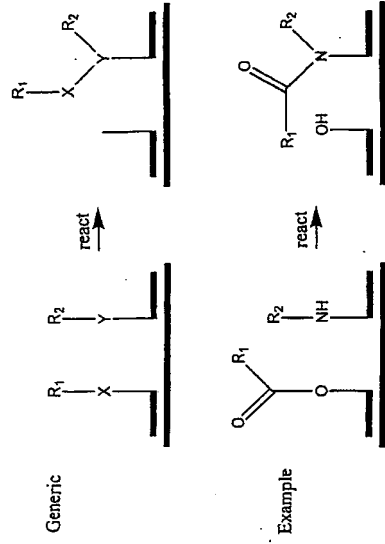


Fig. 2

A. Direct transfer reaction



B. Indirect transfer reaction (reaction followed by cleavage)

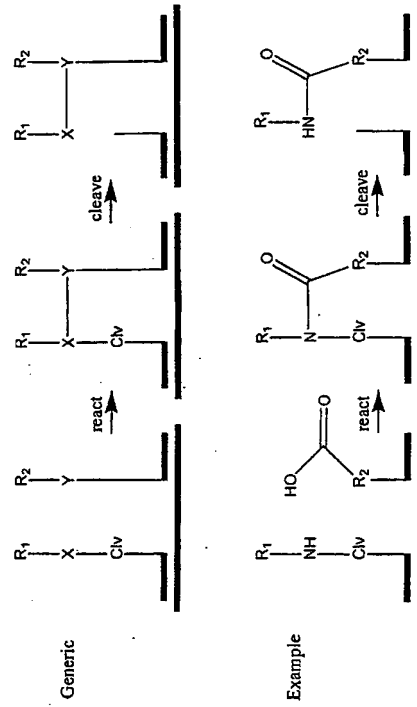
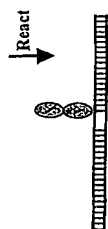
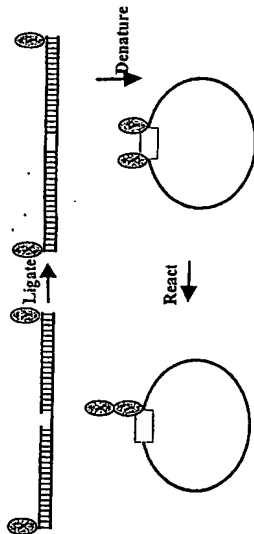


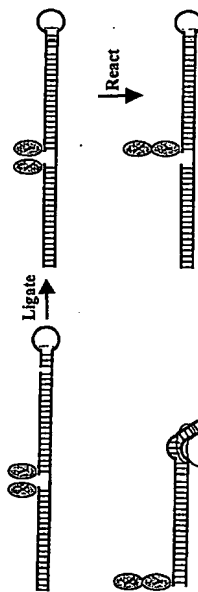
Fig. 3



Example 2:



Example 3:



Example 4:

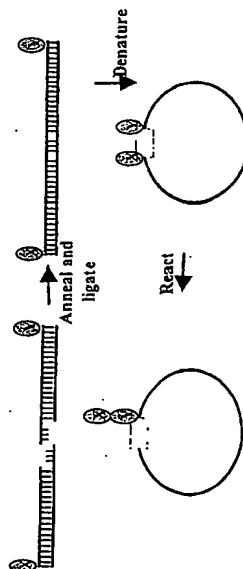


Fig. 4

Example 5:

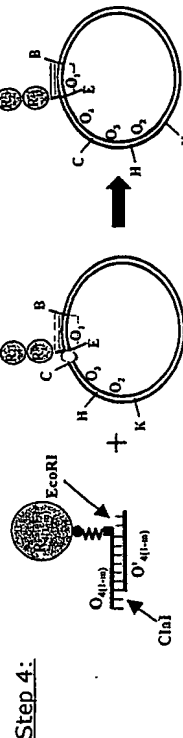
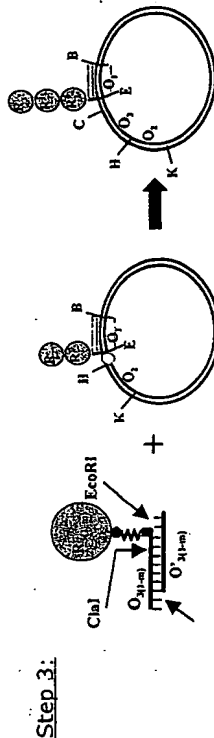
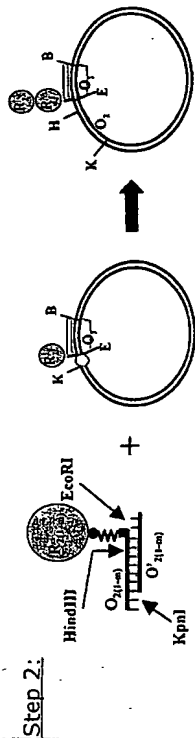
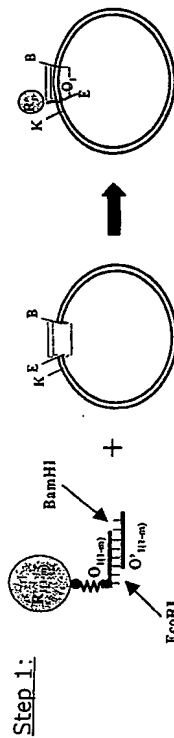


Fig. 4 cont.

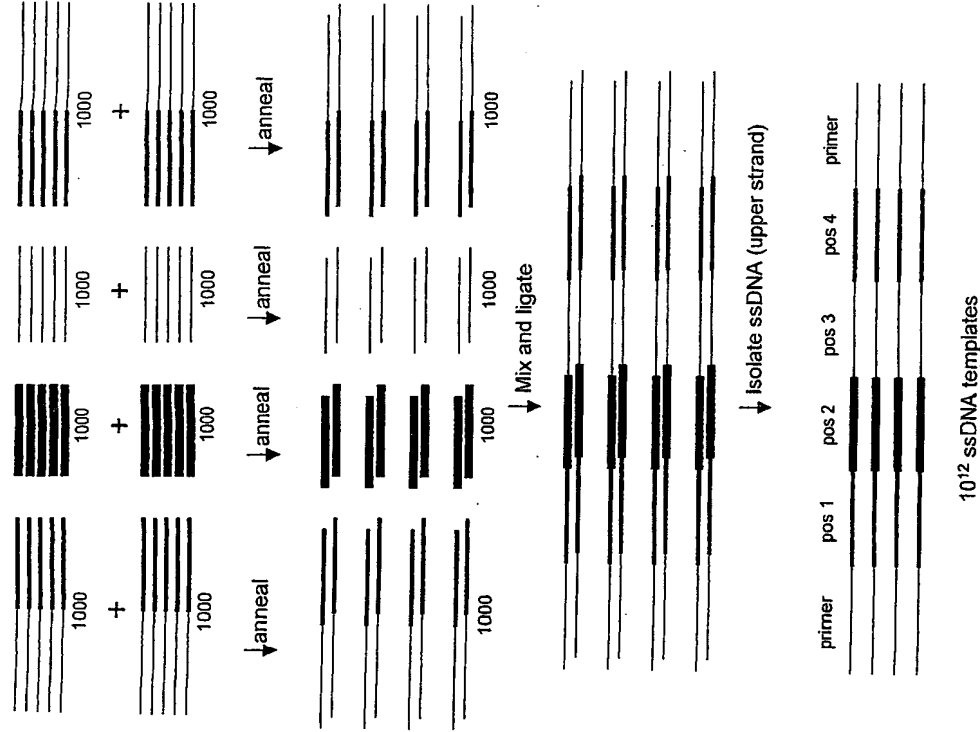
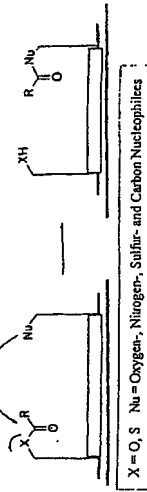


Fig. 5

Nucleophilic substitution by activation of electrophiles

A. Acylation - principle



B. Acylation

Amide formation by reaction of amines with activated esters



C. Acylation

Pyrazolone formation by reaction of hydrazines with β-ketoesters



D. Acylation

Isoxazolone formation by reaction of hydroxyl-amines with β-ketoesters

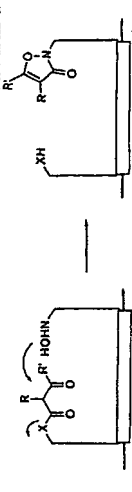
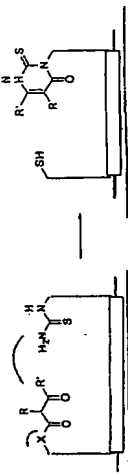


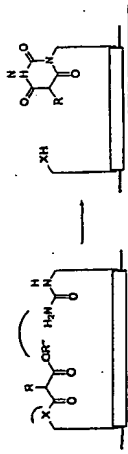
Fig. 6

E. Acylation

Pyrimidine formation by reaction of thioureas with β -Ketoesters

F. Acylation

Pyrimidine formation by reaction of ureas with Malonates



G. Acylation

Coumarine or quinolinone formation by a Heck reaction followed by a nucleophilic substitution



X = O, S X' = Halogen, OTf, OMs Z = O, NH

H. Acylation

Phthalhydrazide formation by reaction of Hydrazines and Phthalimides

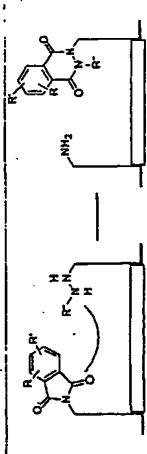
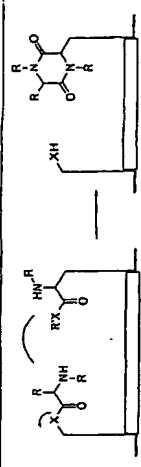


Fig. 6 cont.

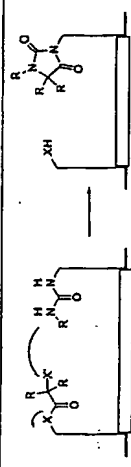
I. Acylation

Diketopiperazine formation by reaction of Amino Acid Esters



X = O, S R' = H, R

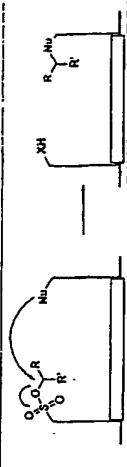
J. Acylation

Hydantoin formation by reaction of Urea and α -substituted Esters

X = O, S X' = Halogen, OTf, OMs, etc.

K. Alkylation - principle

Alkylated compounds by reaction of Sulfonates with Nucleophiles



Nu = Oxygen-, Nitrogen-, Sulfur-, and Carbon Nucleophiles

L. Vinylation - principle



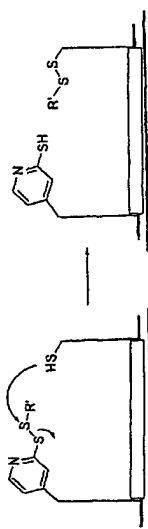
Z = CN, COOR, COR, NO2, SO2R, S(=O)R, SO2NR2, F

Nu = Oxygen-, Nitrogen-, Sulfur- and Carbon Nucleophiles

Fig. 6 cont.

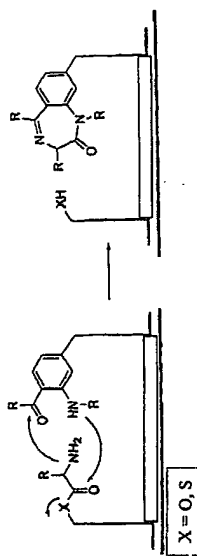
M. Heteroatom electrophiles

Disulfide formation by reaction of Pyridyl disulfide with Mercaptanes



N. Acylation

Benzodiazepinone formation by reaction of Amino Acid Esters and Amino Ketones



Addition to carbon-hetero multiple bonds

O. Wittig/Horner-Wittig-Emmon reagents
Substituted alkene formation by reaction of Phosphonates with Aldehydes or Ketones

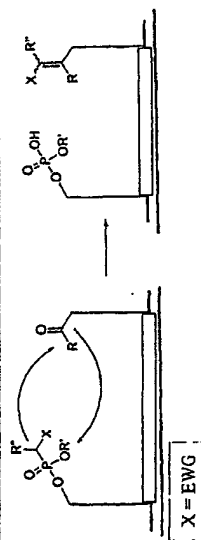
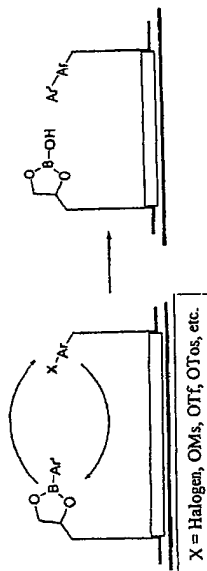


Fig. 6 cont.

Transition metal catalyzed reactions

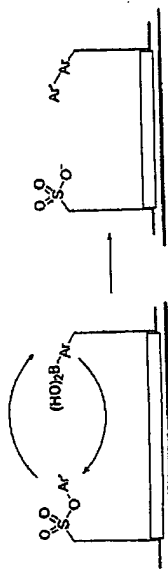
P. Arylation

Biaryl formation by reaction of Boronates with Aryls or Heteroaryls



Q. Arylation

Biaryl formation by reaction of Boronates with Aryls or Heteroaryls



R. Arylation

Vinylarene formation by reaction of alkenes with Aryls or Heteroaryls

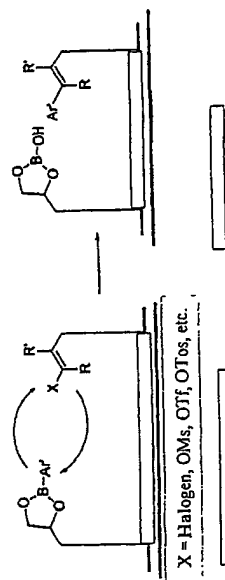
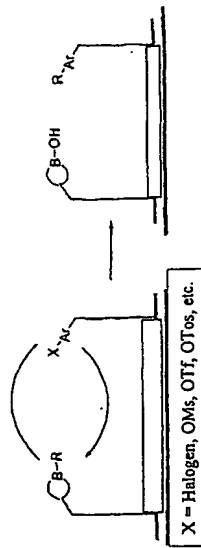


Fig. 6 cont.

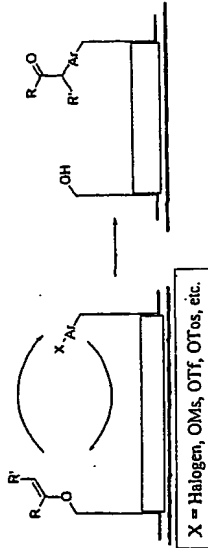
S. Alkylation

Alkylation of arenes/heteroarenes by reaction with Alkyl boronates



T. Alkylation

Alkylation of arenes/heteroarenes by reaction with Enoethers



Nucleophilic substitution using activation of nucleophiles

Condensations

Alkylation of Aldehydes with Enoethers or Enamines

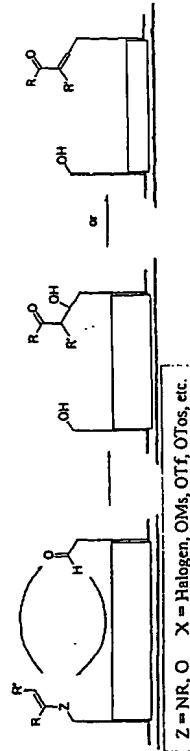
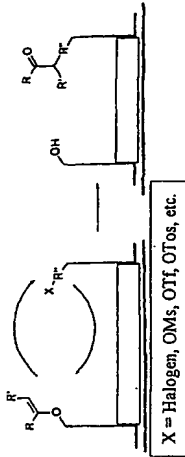


Fig. 6 cont.

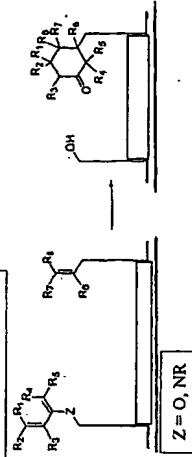
V. Alkylation

Alkylation of aliphatic halides or tosylates with Enoethers or enamines

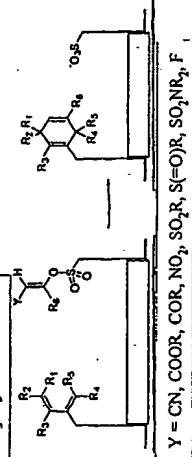


Cycloadditions

W. [2+4] Cycloadditions



X. [2+4] Cycloadditions



Y. [3+2] Cycloadditions

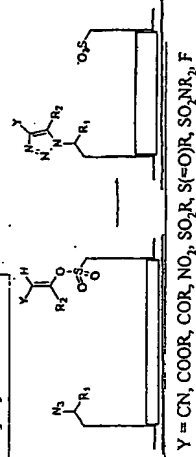


Fig. 6 cont.

Z. [3+2] Cycloadditions

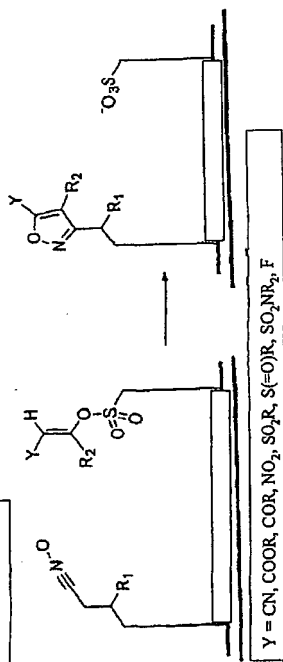
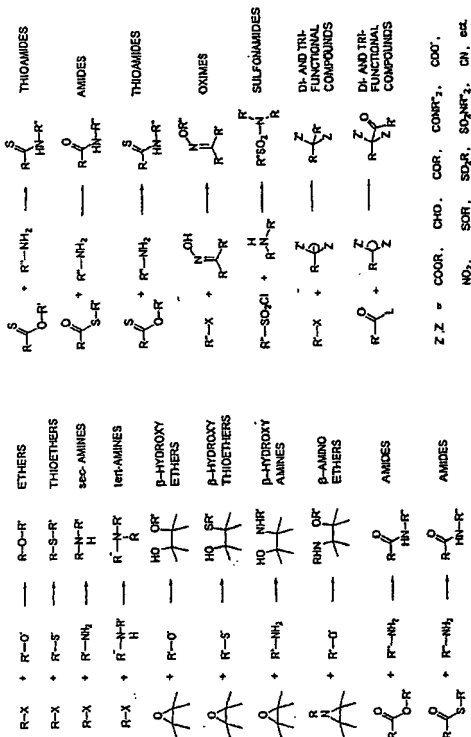


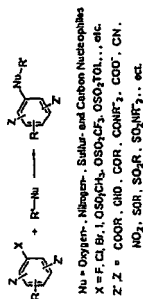
Fig. 6 cont.

Nucleophilic substitution reaction



Aromatic nucleophilic substitution

SUBSTITUTED AROMATIC COMPOUNDS



Transition metal catalysed reactions

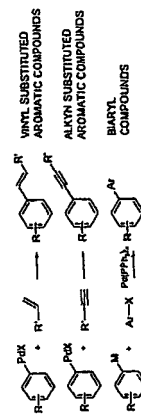


Fig. 7

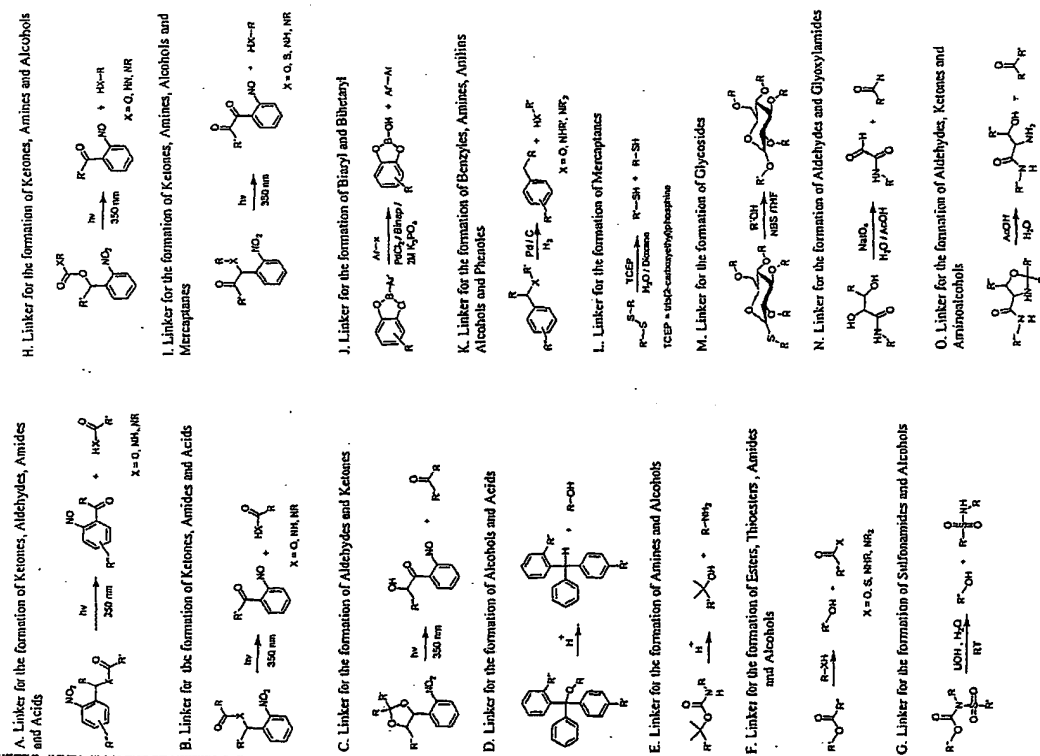


Fig. 8

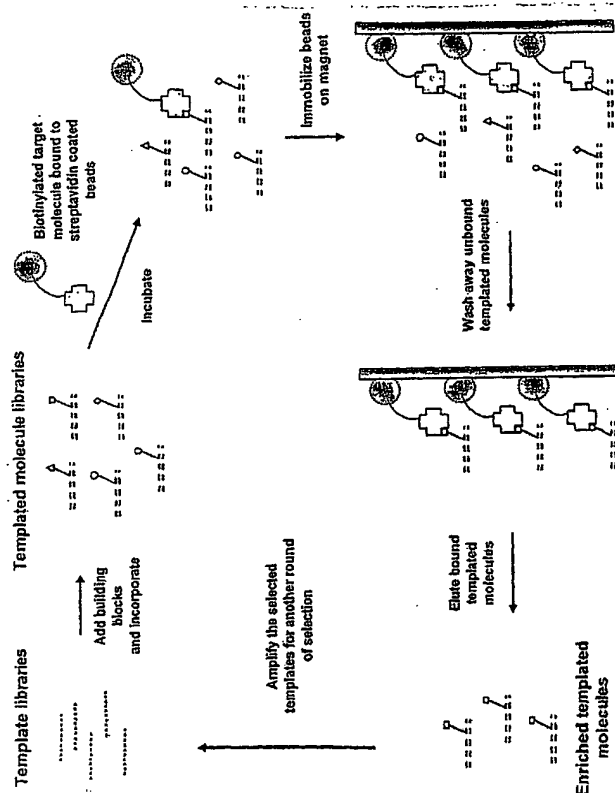


Fig. 9

- In vivo and in vitro amplification, recombination and mutagenesis
- Kunkel site-directed mutagenesis, using one or multiple (e.g., 50) different mutagenic oligos at below-saturating concentrations, i.e., generating a combinatorial library
- USE (Unique Site-directed Elimination), using one or multiple (e.g., 50) different mutagenic oligos at below-saturating concentrations, i.e., generating a combinatorial library
- PCR (Polymerase Chain Reaction)
- LCR (Ligase Chain Reaction)
- PCR shuffling, including family shuffling (shuffling sequences containing blocks with particular homology), and directed shuffling where oligos are spiked into the reaction to direct the shuffling process in a certain direction
- Other types of shuffling, e.g. homologous recombination in yeast; shuffling protocols as developed at the companies Phyllos, Energy Biosystems, Diversa and by Frances Arnold.
- Cassette mutagenesis
- Other polymerase- or PCR-based methods, e.g., overlap extension, gene synthesis, and error-prone PCR
- Chemical or UV-induced mutagenesis
- Wildtype or variant template synthesis and translation into templated polymer (wildtype in this respect means the template sequence that will template the synthesis of the known ("wildtype") polymer, variant in this respect means a partly randomised or spiked template sequence that will template the synthesis of a variant of the known polymer)
- Specific cleavage by restriction enzymes
- Ligation by DNA or RNA ligases; "gene splicing"
- Affinity selections (using the template-templated polymer complex)
- Sequencing
- Arraying the polymers on "DNA chips", by using the template as a tag that binds a DNA array

Fig. 10

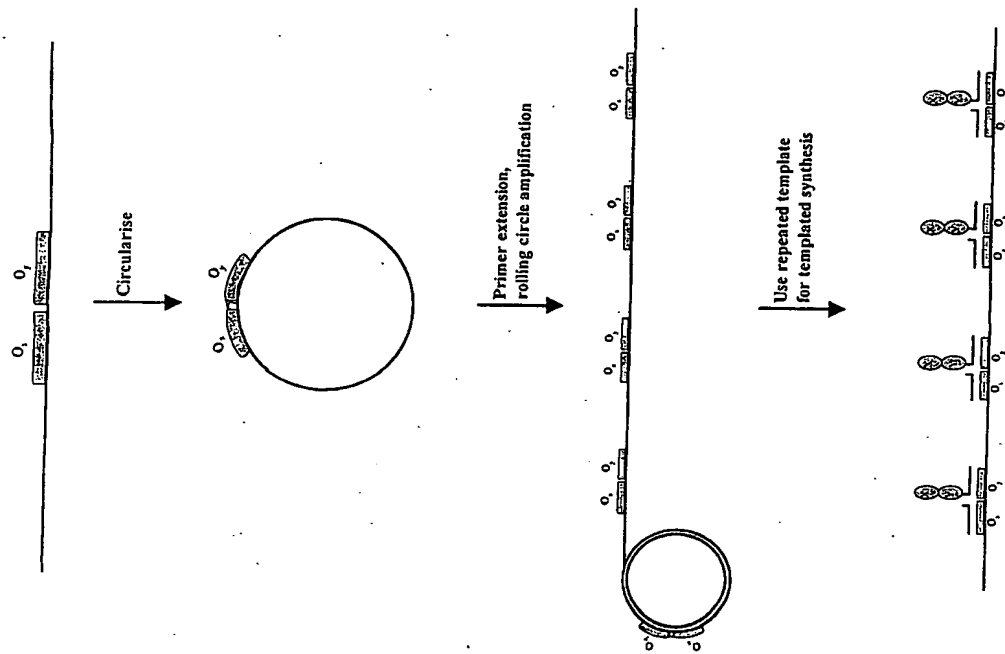
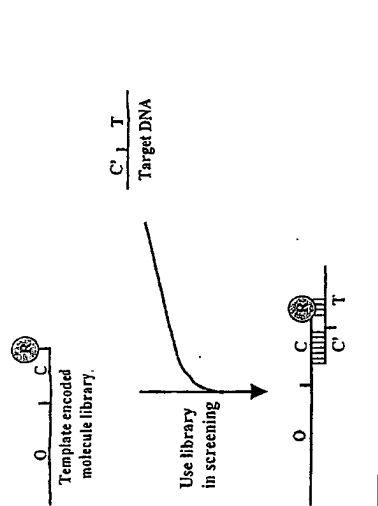
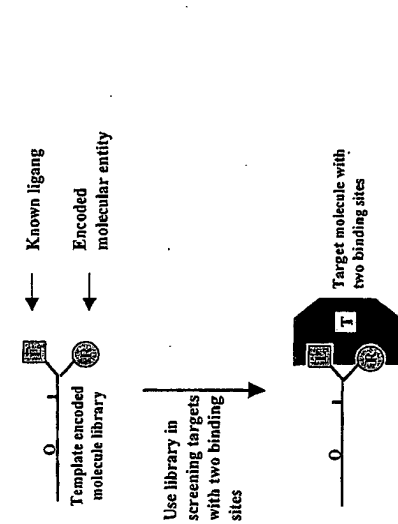
Example A:

Fig. 11



Example E:



Example F:

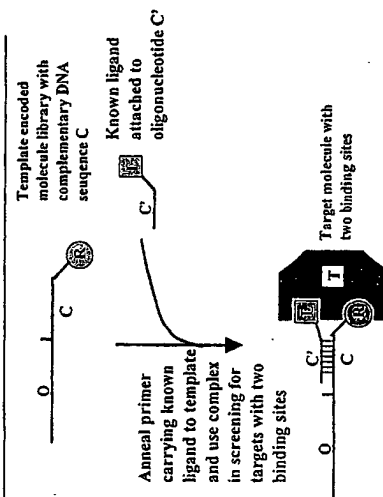
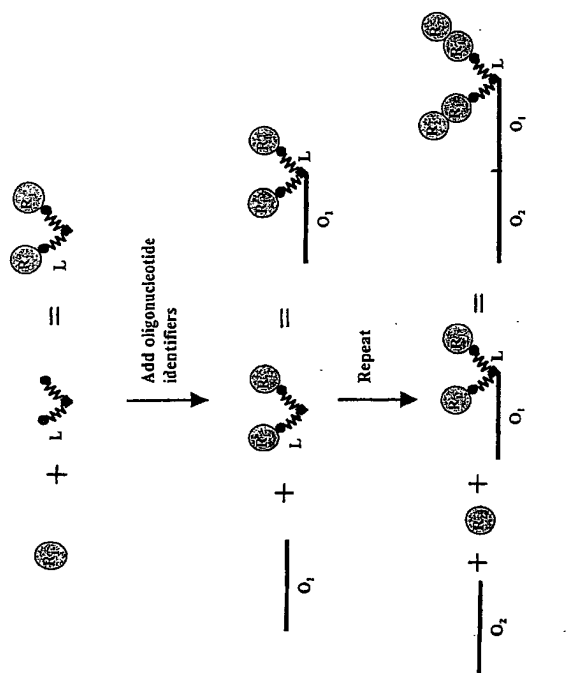


Fig. 11 cont.

Example B:



Example C:

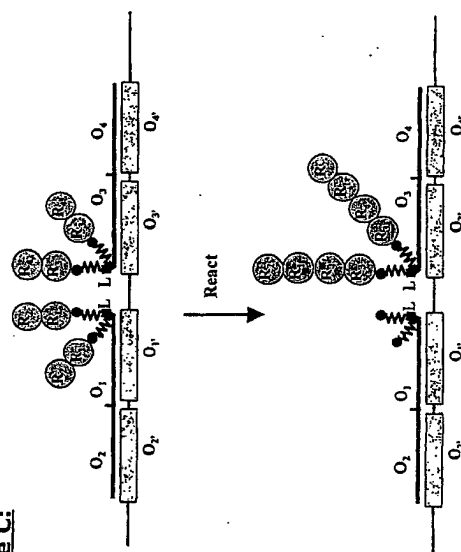


Fig. 11 cont.

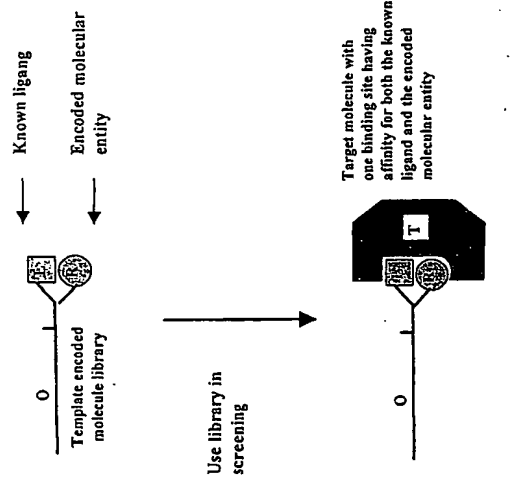
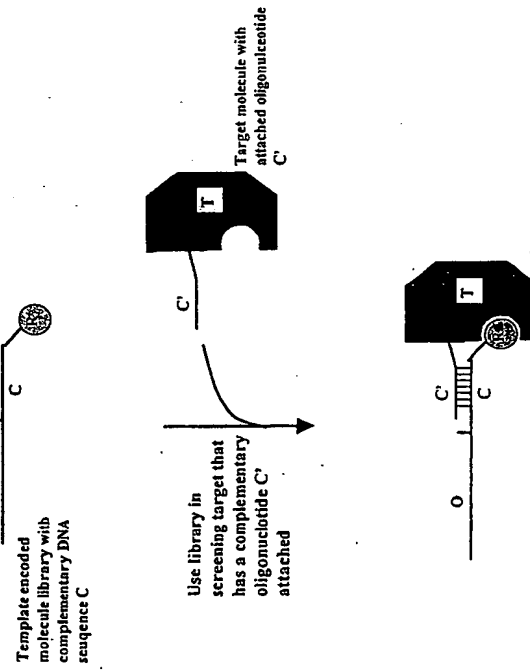
**Example H:**

Fig. 11 cont.

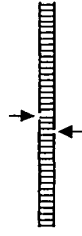
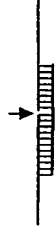
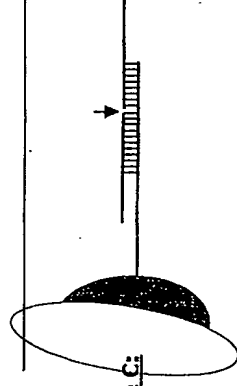
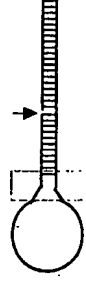
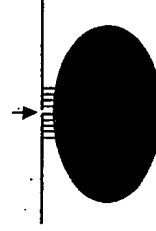
Example A:**Example B:****Example C:****Example D:****Example E:**

Fig. 12

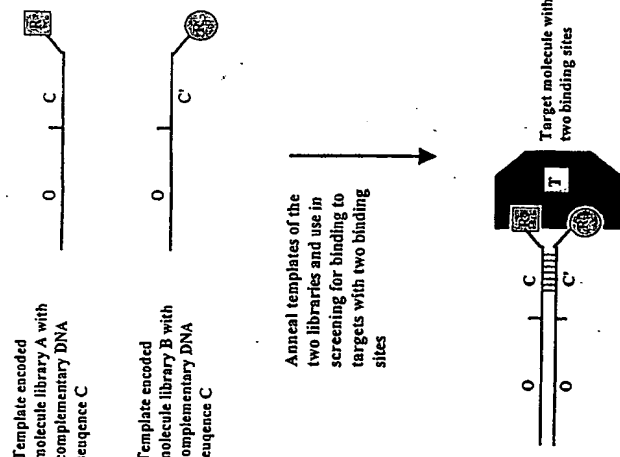


Fig. 14

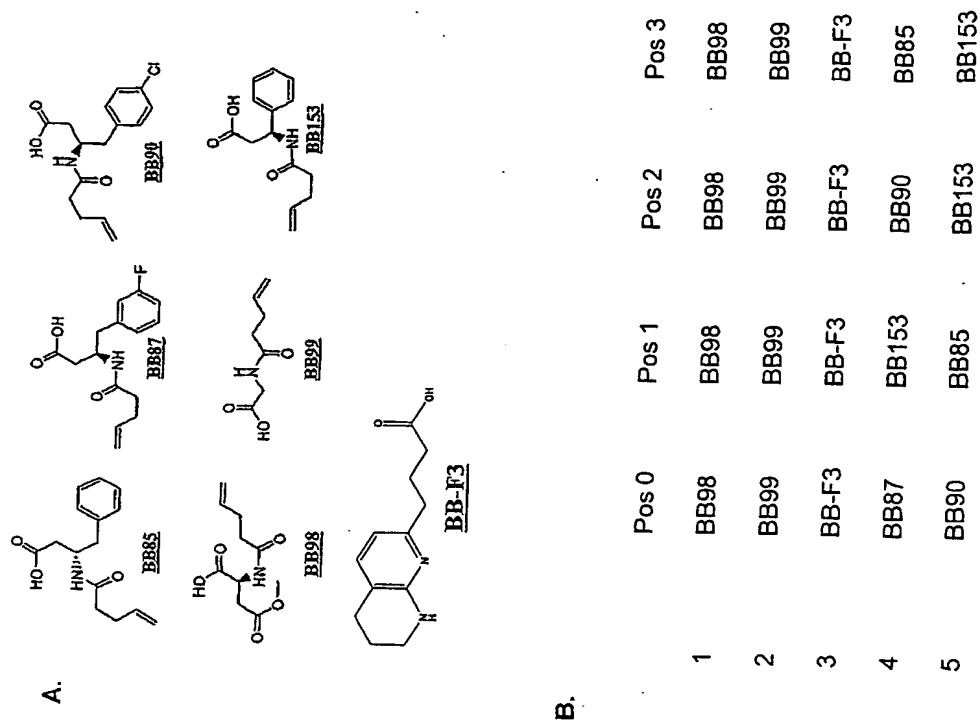


Fig. 15

THIS PAGE BLANK (USPT